Noise, uncertainty, and interest: Predictive coding and cognitive penetration

Jona Vance (Northern Arizona University
Dustin Stokes (University of Utah)[1]

As the various contributions to this special volume will no doubt attest, the predictive coding model of the mind and brain is remarkably ambitious in scope of explanation. For one, its central proponents allege that it offers a single, unified mechanism for phenomena that are traditionally treated as distinct and, accordingly, explained in disparate ways. As Jakob Hohwy writes,

> Perception, action, and attention are but three different ways of doing the very same thing. All three ways must be balanced carefully with each other in order to get the world right. The unity of conscious perception, the nature of the self, and our knowledge of our private mental world is at heart grounded in our attempts to optimize predictions about our ongoing sensory input. (Hohwy 2013: 2).

The mechanism that allegedly unifies these phenomena—prediction error minimization—is also supposed to promise unity across the cognitive sciences. As Andy Clark writes, this framework

> makes rich and illuminating contact with work in cognitive neuroscience while boasting a firm foundation in computational modeling and Bayesian theory. It thus offers what is arguably the first truly systematic bridge linking three of our most promising tools for understanding mind and reason: cognitive neuroscience, computational modelling, and probabilistic Bayesian approaches to dealing with evidence and uncertainty (Clark, 2013 190-1).

Ambition such as this begets excitement, but it should also encourage careful scrutiny. Our general interest is at the interface of two standardly distinguished mental kinds or processes—cognition and sense perception. Both of the above theorists, as well as others, embrace to some degree a radical consequence of the predictive coding framework, namely, that the cognition/perception distinction will have to be revised in some important ways, if not abandoned outright. An important way to test this feature of the framework, as well as use it to shed new insight

[1] This work was thoroughly collaborative and the paper thoroughly co-authored--the order of authors was chosen randomly.

on the cognition/perception distinction, is to consider how the models within the framework analyze and explain a specific, possible cognitive-perceptual relation, namely, the cognitive penetration of perception. That is the topic of this paper: how do extant theorists of predictive coding conceptualize and explain possible instances of cognitive penetration? Jakob Hohwy (2013) has offered the most comprehensive analysis on this set of points, and so the analysis offered here focuses primarily on his discussions.[2]

The paper proceeds as follows. §I offers brief clarification of the predictive coding framework and relevant mechanisms, and a brief characterization of cognitive penetration and some challenges that come with defining it. §II develops more precise ways that the predictive coding framework can explain, and of course thereby allow for, genuine top-down causal effects on perceptual experience, of the kind discussed in the context of cognitive penetration. §III develops these insights further with an eye towards tracking one extant criterion for cognitive penetration, namely, that the relevant cognitive effects on perception must be sufficiently direct. Throughout these discussions, we extend the analyses of the predictive coding models, as we know them. So one open question that surfaces is how much of the extended analyses are genuinely just part of the predictive coding models, or something that must be added to them in order to generate these additional explanatory benefits. In §IV, we analyze and criticize a claim made by some theorists of predictive coding, namely, that (interesting) instances of cognitive penetration tend to occur in perceptual circumstances involving substantial noise or uncertainty. It is here that our analysis is most critical. We recognize, and indeed take pains to carefully identify, why the theorist of predictive coding is motivated to make this claim. However, we argue that, when applied, the claim fails to explain (or perhaps even be consistent with) a large range of important and uncontroversially

[2] We will, when relevant, attempt to connect our analysis with other examples of predictive coding theorizing. But again, beyond Hohwy 2013, there are few if any extended discussions of predictive coding *and* cognitive penetration. (And this we take it to be one of the virtues of this special issue: there will hopefully be further analysis of this interface.)

interesting possible cases of cognitive penetration. We conclude with a general speculation about how the recent work on the predictive mind may influence the current dialectic concerning top-down effects on perception.

I. Predictive coding and cognitive penetrability

Broadly, predictive coding theories embrace an inferential approach to perception, which can be traced back at least to Helmholtz (1867). And Clark (2013) traces the idea of giving a central role for error signals in all brain processes to the influential cybernetician W. Ross Ashby, who claimed that "the whole function of the brain is summed up in: error correction" (Ashby 1947). Clark interprets this provocative claim with a provocative claim of his own: "brains are essentially prediction machines" (Clark 2013).

More precisely, predictive coding can be thought of as a framework within which one can develop specific, testable models of mental processing. The predictive coding framework is primarily distinguished from other frameworks by the following core commitment: a single mechanism of prediction error minimization plays a central role in all or most mental processes. Accordingly, let's refer to the overall framework as 'PCF' (predictive coding framework) and the central mechanism as 'PEM' (prediction error minimization).

Within the PCF, the most prominent extant proposals for how to develop models make further commitments about the processing that carries out prediction error minimization. These commitments entail that the processing is probabilistic, hierarchical, and dynamic. We illustrate these further commitments with respect to perceptual processing. Begin with the standard assumption that the objects, events, and features in the environment stimulate sensory receptors of a given perceiver. According to standard developments of PCF, perceptual processing draws on stored representations of the distal environment approximating a probability density function ("priors") and predictions concerning which sensory inputs are likely given the priors ("likelihoods"). So for any given hypothesis h and some bottom-up sensory information e, there are two probabilities of initial relevance: the prior probability of h independent of the current evidence, $P(h)$, and the likelihood, which is the probability of the occurrence of e given h, $P(e|h)$. Given Bayes' theorem, these two probabilities can be combined to yield the probability of h given e, $P(h|e)$, the posterior probability.

According to standard, extant PCF accounts, the PEM process (which subsumes perceptual processes) is implemented across a hierarchy of layers of nodes. In humans, nodes are realized by neural populations, but each node in the model need not correspond to a single neuron. The lowest layer of the hierarchy can be identified with external sensory receptors or areas of processing close to the sensory periphery (e.g. the lateral geniculate nucleus in Lee & Mumford 2003: 1435). There are then a number of other layers above the lowest layer. At each layer, states predict the activity at the layer immediately below. So, for example, states at layer n, aim to predict activity at layer n-1. And states at n-1 aim to predict activity at layer n-2, and so on down to the lowest layer. Each level serves as the prior with respect to activity at the layer below. When the brain correctly anticipates the activity at the layer below, those aspects of the sensory data are "explained away". When the brain doesn't anticipate aspects of the data, prediction error occurs. Prediction errors are the remains of the sensory data as it is sent up through the neural pathways of the hierarchy. They constitute what is left for higher layers to explain away. Together these top-down and bottom-up information processes constitute what is sometimes called 'message-passing' (Yedidia, Freeman, & Weiss 2001). That the bottom-up signal is an error signal is what distinguishes predictive coding approaches from all others.

These are the most basic features of the PCF, and its reliance on PEM. And again it should be recalled that the central advocates of this view argue that this mechanism is at the foundation of, and thus explains, not just so-called higher level mental processes like belief formation and reasoning, but all of perceptual representation, and additionally attentional processes, action, and even the agential sense of self. Invoking once more Clark's slogan, PCF says that the brain is essentially a prediction machine.

Cognitive penetration of perception (CP) is a possible phenomenon the reality of which is much debated. And so accordingly, it is much less straightforward to characterize, and partly because it is the definition of the phenomenon that is partly under dispute. The general idea is this: do mental states like belief or intention or desire—'propositional attitudes' in the terms of the philosopher— influence sensory perception *in some important or interesting way*?[3] This is sometimes put in terms of processing: does information in the former processes influence the computational processing of sensory systems? Sometimes it is put in terms of phenomenal experience: does one's belief somehow importantly affect the look or sound or feel of things? Either way, if the answer is 'yes', then perception is supposed to be cognitively penetrable. If 'no', then perception is cognitively impenetrable. A question that quickly surfaces, and that will occupy much of the analysis in this paper, concerns the use of 'interesting': what kinds of cognitive effects on perception count as interesting or important?

One way to characterize the importance of the relevant class of cognitive effects on perception is in terms of the kinds of consequences that the phenomenon would have, should it actually occur. These consequences typically divide into two kinds: epistemic consequences and scientific-theoretical consequences. Perception is supposed to provide some objective representation

[3] As an anonymous reviewer helpfully points out, although the standard emphasis in the literature is on penetration by propositional attitudes, especially belief, one may consider other possible top-down influences. One possibility is influence on perception by desires and emotions. For related philosophical discussion, see Siegel 2012, Stokes 2012;. And there is certainly suggestive research in perceptual psychology. For a few recent examples, see Cole et al 2013; Stefanucci et al 2012; Stefanucci et al 2011; Stefanucci 2010; Stefanucci and Proffitt 2009. In spite of this important caveat, and area for new philosophical research, our focus will largely be on belief as the relevant candidate penetrating state. As will become clear, this is for the simple reason that belief is the most likely candidate penetrating state when working from within the PCF (albeit, belief is construed more liberally on this conceptual framework).

of one's immediate environment and, thereby, provide a foundation for empirical knowledge. But, if perception is influenced partly by background beliefs or other cognitive attitudes, then the objectivity of perceptual representation might be questioned. So, one way that a type of cognitive effect on perception could be important or interesting in the relevant way, is if it implies some substantive consequence for the epistemology of perception.[4] The second consequence concerns how the architecture of the mind is theorized. Cognition and sensory perception are distinguished by a number of theorists, and it is often supposed that sensory perception operates in a way that is largely, functionally independent of cognitive states and processes like belief, desire, and intention. The most robust version of this thesis is the strong modularity theory of mind, which maintains that sensory systems are *informationally encapsulated* from cognitive systems (Fodor 1983; Pylyshyn 1999). Relevant cognitive effects on perception, then, would be ones that force revision of these empirical models of the mind.[5]

A (perhaps) alternative way that the importance or interest-value of the relevant cognitive effects on perception has been put concerns the directness of the cognitive-perceptual relation. Put in its strongest form, cognitive penetration occurs only if sensory systems compute in a way that draws directly on information stored in cognitive systems. Raftopoulos and Zeimbekis attribute this *vehicle-based condition* to Zenon Pylyshyn (1999), where Pylyshyn seems to employ it to exclude cases where cognition drives overt shifts in spatial attention (thus, some external action), which in turn causes changes in sensory processing (Raftopoulos and Zeimbekis 2015: 27-32).[6] Although typically not made explicit, one way to motivate the directness criterion is by appeal to the consequences just

[4] For more discussion of the epistemology of cognitive penetration of perception, see Lyons 2011, Siegel 2012, Silins forthcoming, Vance 2014.

[5] For a complete analysis of these consequences, including how they relate, and how they might be used to define or characterize cognitive penetration, see Stokes 2015.

[6] Fodor 1988 makes very similar claims, where attention-shift cases seem to be excluded by virtue of failing to involve a direct cognitive effect on perception.

discussed. If a cognitive effect on perception takes a rather indirect or circuitous route—for example through a series of intentional actions—then such a cognitive-perceptual relation implies nothing particularly important about the epistemology or architecture *of perception*. Thus, a theorist like Pylyshyn might conclude, cognitive penetration (as a theoretically important cognitive-perceptual phenomenon) requires a direct cognitive effect on perception.

There are reasons to be cautious about the directness condition on cognitive penetration.[7] But supposing that one commits to it as necessary, then interestingly the predictive coding framework provides one clear way to see how cognition could directly affect perception. Put most simply, and to be developed in the section that follows, on this model prior expectations (at least some of which would count by all lights as prior *beliefs*) directly influence how sensory input is processed and across the many layers of neural architecture. In at least one respect, this mirrors the project of Macpherson (2012), who provides a nomologically possible model of *indirect* cognitive influences on perception that might nonetheless count as cognitive penetration. The application of PCF here generates a nomologically possible model of *direct* and pervasive cognitive influence on perception.

II *Top-down prediction versus top-down causal influences*

In *The Predictive Mind*, Hohwy considers two arguments regarding the relation between PCF and cognitive penetrability of perception. The first, which we will consider here, is an argument that concludes that PCF implies that cognitive penetration should occur relatively frequently. The second, which we will discuss in section IV below, concludes that cognitive penetration does not occur (at least from within the PCF). We are less interested in the success of these arguments, and

[7] See Macpherson 2012; Stokes 2012, unpublished MS; Raftopoulos and Zeimbekis 2015.

more in how analysis of various premises therein illuminate both further details of and potential problems for the PCF. The first argument runs as follows:

> 1. Perceptual inference always depends on top-down prediction of sensory input.
> 2. There is no theoretical or anatomical border preventing top-down projections from high to low levels of the perceptual hierarchy.
> 3. Cognitive penetrability occurs when the influence of sensory input is changed by top-down input from high to low levels of processing.
> 4. So, cognitive penetrability should be expected (Hohwhy 2013: 122).

PCF entails that the prediction process can have cascading effects down through the hierarchy. However, from standard sketches of the framework (as discussed in I. above) it's not clear why. The standard sketch focuses on top-down *prediction*, and prediction alone does not entail any causal influence. Compare: if you predict rain tomorrow, your prediction does not thereby have any causal influence on whether it will rain tomorrow. For all the sketch says, the system could simply predict the activity at the layer below without having any effect on the activity at the layer below. If that were the case, then there wouldn't be any top-down *influence* on the layers below; there would only be top-down predictions used to generate error signals sent back up the hierarchy. Put in the terms of Hohwy's argument above, the PCF needs to be enhanced with a story about top-down causal influence, else premises 1 and 2 simply do not hook up with premise 3 to motivate the conclusion. Instead, the most we can plausibly conclude, given a premise about prediction (P1) and a premise about consistency between the model and top-down projections (P2), is a rather weak conclusion that PCF is compatible with cognitive penetrability. What follows is an attempt to fill in this gap, which better motivates Hohwy's conclusion in the above argument and, more generally, offers insight on how top-down influences on perception should be theorized from within the PCF.

What's needed is an explanation of why PCF entails that there can be top-down influence by one layer on the layer below and on subsequent layers as well, rather than just downward prediction. To do so, it is important to understand that the prediction process can be described in terms of the *activation* of nodes in a belief net. Once we have the basics of the top-down influence described for simple belief nets, we can generalize for more complex nets, including the brain. Consider a simple belief net with three layers: top, middle, and bottom (Fig. 1).[8] The top and middle layers are hidden layers. The bottom layer is the visible layer. In artificial neural networks, the visible layer receives training data and the hidden layers are there to try to reconstruct the training data to build an accurate model. The solid downward arrows indicate the generative weights. They encode the system's representational model of the world. The upward dashed arrows indicate the recognition weights. Suppose the top layer is activated--that is, the nodes/neurons at the top layer turn on or off in some combination. In the simple net, nodes in the network can be activated in either of two phases. There's a bottom up 'wake' phase and a top down 'sleep' phase (Hinton et al. 1995). In the sleep phase, when the top layer is activated in a certain way, the generative weights between the top and middle layer cause a certain pattern of activation at the middle layer. In the wake phase, the activation at the middle layer causes the upward recognition weights to engage and send information to the top layer. This causes a new pattern of activation at the top layer, which is one way to initiate the generative weights again, and so on. In "predicting" the activity at the layer below, the generative weights cause activation at the lower layer, which aims at reconstructing the actual activity at the lower layer. So the "prediction" is more than just a prediction about the layer below: it has a causal impact on activation at the layer below. That's how "predictions" can have a top-down, causal effect on lower level activity. Nonetheless, it's a prediction because, in activating these lower neurons, the system is trying to accurately reconstruct the actual activity at the lower layer; it is trying to predict which states are activated at that layer (either typically or at a particular time).[9]

---

[8] Note that there are important differences between this simple net and the mental architecture hypothesized in PCF. We discuss some of these differences below.

[9] See Vance, in preparation, for more an extension of these points to processes resulting in action in PCF.

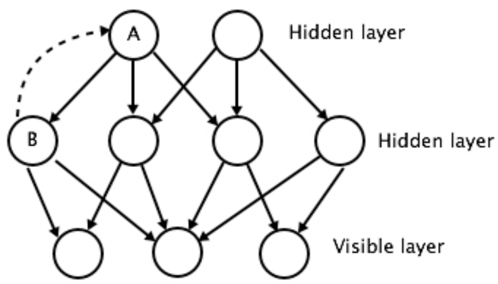

Fig. 1 A belief net with three layers. The bottom layer of nodes constitutes the visible layer of the network; the middle and top layers are hidden layers. The downward solid arrows indicate generative weights connecting adjacent layers of the network. When the top layer is activated in the 'sleep phase' (i.e. the nodes are turned on in various patterns), this initiates a message passed by the downward generative weights and causes them to activate nodes at the middle layer. The upward solid arrow represents a recognition weight. When node B is activated in the 'wake phase', this connection passes information to the top layer.

Generalizing, consider a net with n layers. Suppose that neurons at the top layer n are activated in a certain pattern. Given the particular pattern of activation at layer n, the top-down generative weights activate the neurons at the layer immediately below (n-1) in a prescribed pattern as well. The generative process can be repeated downward through the hierarchy. Once the neurons at n-1 are activated because of the influence from the activation at layer n and the generative weights between n and n-1, the generative weights connecting layers n-1 to n-2 are engaged by the new activation at n-1 and cause a prescribed pattern of activation at layer n-2. Once the activation at n-2 occurs, the generative weights between n-2 and n-3 engage, and so on down to the layer just above the lowest level of input data--which is the initial sensory data received from the receptors.

According to Hohwy (2013), for at least some of the layers in the hierarchy (presumably relatively low layers, near the sensory receptors), the activation at each of those layers helps gives rise to conscious perceptual experience. So the top-down influence by activation of higher levels, guided by the generative weights, can have an effect on activation at lower levels, which can make a difference to what gets consciously experienced in each perceptual modality. It is in this way that a theorist of the PCF *might* conclude with some force that if the mind centrally proceeds according to a prediction error mechanism, then the cognitive penetration of perception (at the conscious experiential level) should be expected to occur more often than more traditional models or frameworks suggest. The open question that remains is how much of the causal story just offered is in fact part of that very mechanism, or something over and above it.

III *Directness*

It is uncontroversial that cognition can have some influence on perception. For example, suppose that one desires cookies and believes that there are cookies in the pantry. Suppose that this causes one to walk to the pantry and look inside, and one then sees cookies. This does not count as cognitive penetration. The influence from cognition to perception proceeded via changes in action which in turn caused changes in impacts on the subject's sensory receptors.

In response to the above kind of case, it is natural to propose a directness criterion for cognitive penetration. That is, it's natural to think that one thing that precludes the above case from being genuine CP is that the process by which cognition had an effect on perception was not appropriately or sufficiently direct. We now explore how one could formulate directness criteria using PCF.

One way to characterize direct influence appeals to relationships between nodes in belief nets or graphical models more generally. It says that some node, A, has direct influence on another node, B, if and only if A has influence on B even after we account for all the other influences on B (Danks 2014: Ch 3).[10] In graphical terms, the only direct influences are those between adjacent nodes; such as A and B in Fig. 1. Applying this conception of directness to cognitive penetration yields the following criterion: cognitive penetration occurs only if some cognitive state is realized by

[10] Danks does not offer this characterization of directness in relation to cognitive penetration.

a cognitive node C and some perceptual state is realized by node P, C has an influence on P, and C and P are adjacent.

The above proposal faces at least two difficulties. First, the proposal implies that cognitive penetration occurs only if there is a cognitive node C adjacent to a perceptual node P. This requires a controversially sharp boundary between cognitive and perceptual layers. Suppose that there are clear cases of cognitive nodes and clear cases of perceptual nodes, but vague layers in between. For example, there is a cognitive node C adjacent to a vaguely categorized node V, and V is adjacent to a perceptual node P. And suppose that C has direct influence on V, which has direct influence on P. Such a pathway would not allow for cognitive penetration as defined above, since there is no pair of adjacent nodes where one is cognitive and the other perceptual. Nevertheless, top-down influence along a pathway from C to V to P seems a plausible candidate for cognitive penetration. Second, the proposal identifies a kind of directness more restrictive than any we are aware of. Extant appeals to a directness requirement for cognitive penetration seem to be motivated to rule out causal chains that are indirect in the sense that they deviate from internal mental processes (Pylyshyn 1999; Raftopoulos & Zeimbekis 2015). An argument would be required to motivate this more restrictive criterion, and we know of no such argument.

A perhaps more promising way to characterize the notion of directness concerns the kinds of information carrying vehicles or states involved. As Raftopoulos and Zeimbekis describe such a criterion, "…for there to be CP, the perceptual system has to draw directly on the informational resources of a cognitive system in performing its computations. This condition is implicit in Pylyshyn's denial that attentionally mediated cognitive influences on perception qualify as CP (1999: 344; 2003: 90)" (Raftopoulos and Zeimbekis 2015: 27).

Although the notion of a vehicle is not entirely clear, the top-down generative processes described in PCF plausibly could meet the proposed "vehicle" criterion. Within this framework, the vehicles are the generative weights that cause activation at layers immediately below as a result of activity immediately above. For example, suppose that n is a cognitive layer (we discuss below what a 'cognitive layer' might be in PCF). And suppose n-k is a perceptual layer. And suppose that once there's activity at layer n, the generative weights cause activity at layer n-1, then n-2, and so on to n-k. The generative weights transmit information down from the cognitive layer n to the perceptual layer n-k; they serve as the vehicle of penetration. However, the details of what is being computed at the various layers might well matter here. For example, if layer n-5 were a layer at which the subject generates activity in imagination, that would count on Macpherson's proposal as an "indirect" process (Macpherson 2012). So, the formal point that there are top-down generative weights that can serve as the vehicles by which cognition influences perception is not enough to make it clear that the process of influence is direct in the relevant ways one might want.

A further approach to characterizing directness does so negatively, by specifying the kinds of causal influence through which CP does not occur. For example, one can specify that CP cannot occur via shifts in the impacts on sensory receptors, as in the cookie case above. Here we note that shifts in the impacts on sensory receptors are not required for the top-down influence by the generative weights. So top-down message passing, as described here, satisfies the directness criterion formulated in terms of the lack of change in sensory impact. Similarly, there is nothing built into the top-down message passing process that requires other kinds of changes one might think are incompatible with CP. For example, nothing in this process requires any shifts in the subject's spatial attention. And as some readers will note, a common rejoinder to some alleged cases of CP is to explain the phenomenon in terms of cognitively driven shifts in spatial attention. This explanation—sometimes called the *attention-shift interpretation*—maintains that whatever else might be required of CP, a negative directness condition is that it cannot involve changes in the spatial array of attention, even if those changes result in perceptual differences.[11] The message passing account just given would seem to evade this rejoinder.

We now add a further wrinkle to the account and discuss its relevance for whether CP can occur in PCF. We mentioned above that the top-down influence by the generative weights is usually thought to occur concurrently with the bottom up influence by the recognition weights and the error signal they carry. Suppose that there is a particular pattern of activation at layer n and that that causes activation at n-1 via the generative weights. The activation at n-1 aims to reconstruct or match the actual activation at n-1. That's the sense in which it is a prediction. To the extent that the top-down activation fails to match the actual activation at n-1, an error signal is generated and sent back up to n. The error signal can be sent back up even while the system also sends further top-down messages from n-1 to n-2 to n-3 and so forth. In short, signals can be constantly passed both up and down the hierarchy. As a result, there may not be a clean sweep of messages from top to bottom, without changes in activity at each level that are also due to bottom up signals.

Now, there are least two reasons that the constant bottom-up influences do not prevent the top-down influences from counting as direct. First, the kind of changes to the sensory signal that are supposed to be inconsistent with CP are not changes internal to perceptual processing itself. The very idea of CP can be cast in terms of changes to the internal signal processing due to a certain kind of influence from cognition. So, even if changes to the error signal had to play a causal role in the relevant process, that would not be a departure from the strictest notions of CP. Second, recall that for the simple belief net in Fig. 1, there are two phases to the message passing process: a bottom-up "wake" phase and a top-down "sleep" phase. Distinguishing these two phases allows us to isolate the causal influence carried by each type of process. That is, we can isolate the contribution from the generative top-down process, even if in practice the two processes interact. The isolation is not

[11] For general discussion of the attention-shift interpretation, see Macpherson 2012; Stokes 2012.

merely conceptual in standard presentations of the predictive coding architecture; it is functional and even anatomical in some cases. In PCF, it is often assumed that bottom-up error messages are carried by superficial pyramidal cells and top-down prediction messages are carried by deep pyramidal cells (Egner, Monti, & Summerfeld 2010; Friston 2005, 2009; Mumford 1992). In addition, some functional separation between top-down and bottom-up processes is required by the framework (Clark 2013: 188). Because we can functionally isolate the top-down generative process from the bottom-up error process, we can see a direct pathway for top-down influence.

IV. Noise, uncertainty, and the value of interest

The previous two sections attempted to clarify and extend the claim that cognitive penetration is a common phenomenon if the mind is as PCF theorizes it. We now turn to the opposite set of considerations, namely, reasons for thinking that PCF implies that cognitive penetration does not occur. Defenders of PCF have offered this suggestion for different reasons. For example, Andy Clark has suggested that on the PCF, standard distinctions between perceptual processes and cognitive processes may dissolve (Clark 2013). Plausibly, cognitive penetration of perception presupposes a distinction between perceptual and cognitive processes, even if there is vagueness about exactly where the boundary lies (though see Vetter & Newen 2014 for an opposing view). Accordingly, an implication of Clark's suggestion may be that cognitive penetration is not a genuine phenomenon, or that it is at least conceptually problematic as currently understood.

Hohwy considers an argument to a similar conclusion but grounded in different considerations. The argument runs as follows.

> 1. Perceptual experience is a matter of how top-down predictions suppress prediction error in perceptual and active inference.
> 2. Hence, if there is a difference in the expectations that drive perception, then there must be a difference in the prediction error, that is, a difference in the bottom-up input.

3. Different expectations and similar bottom-up input are needed for cognitive penetration.
4. So, there can be no cognitive penetration (Hohwy 2013: 122-3).

Hohwy wishes to avoid this conclusion, motivated instead to maintain the compatibility of PCF and CP. He attempts to deflect the argument by resisting premises 2 and 3.

Premise 3 can be interpreted in more than one way. As we argued in section III, CP is consistent with some cognitively induced changes to perceptual processing. If premise 3 entails that cognitive penetration is inconsistent with any change in the bottom-up signal as part of perceptual processing, premise 3 is implausible. However, Hohwy may not have this implausible version of premise 3 in mind.

Hohwy suggests in more than one place that the definition of cognitive penetration should be relaxed. He writes, "we should accept the in principle possibility of interesting cases of penetrability even if there is some degree of difference in the sensory input of the cases we are comparing" (Hohwy 2013: 120; see also p.123). Hohwy is here suggesting that we should not define cognitive penetration so as to rule out a case where a subject's cognitive states (say, what she desires and believes) change her perceptual experience *through* the mediation of shifts in spatial attention (or some other type of overt, bodily action). Put in terms familiar to the cognitive penetrability debates, Hohwy proposes that the spatial attention-shift interpretation should not be an explanation that is *alternative* to and therefore *incompatible* with a cognitive penetration interpretation. Hohwy's line of thought here seems to be that if we do restrict the definition in this way, we lose the theoretical benefit of cases that might be instructive or interesting for a theory of cognition and perception.

It is unclear what, more precisely, constitutes "being interesting" in Hohwy's discussion. One way to characterize what makes possible cases of cognitive penetration interesting traces back to those who were the first to discuss it, including those who deny that it happens. As mentioned in §I, modularity theorists maintain that perceptual systems operate in a way that is functionally independent of cognitive systems like belief and goal-orientation. One way an effect on perception could be interesting--*theoretically interesting*--is if it threatened the alleged modularity of perceptual systems. Another, related, way that an effect could be interesting—*epistemically interesting*—is if it implied consequences for the reliability or knowledge-providing role of perception. If interest-value is determined in either or both of these ways, then one could plausibly worry that this is precisely why the definition of cognitive penetration should not be relaxed as Hohwy suggests. A case involving change to the sensory input bears no import to the alleged modularity or epistemic role *of perceptual systems*.[12] And this is true even if change in input is caused by the perceiver's cognitive states and results in experiential changes. So unless Hohwy understands the relevant interest-value in some way other than either of these ways, then his resistance to premise 3 looks unmotivated.[13]

Hohwy's second line of response to the argument above is to resist premise 2. This response is worthy of extended discussion, since it both better illuminates how the theorist of PCF will plausibly analyze cognitive penetration and, relatedly, reveals some deeper problems for the

[12] An analogy here may help. Maintaining for the moment the input-output description, suppose one wants to ask about the value of some process-type. Often the value of a process-type is determined by its output in relation to some target or broader desired outcome. So, consider a familiar issue for process reliabilism in contemporary epistemology. When one asks whether some belief-forming process-type *P* is reliable, one is asking whether *P* delivers output, with high long-run frequency, that veridically represents the world (i.e. *P* frequently produces true beliefs). To sensibly ask this question, one does so in a way conditionalized on input that also stands in that same relation with the world: one assumes that we have good (true) input. This is for the simple reason that we learn nothing about the relevant value of the process (in this case, how frequently it delivers output that veridically represents the target) if we run the process on information that fails to stand in that same relation with the target. Accordingly, bad (false) input to a belief-forming process does not impugn the reliability of the process. Worse, it stifles our attempt to determine the value of that process. As one might recognize, this is the very reason that we ask questions about the validity of arguments conditionalized on true premises. Here the value is truth-conducivity, and an argument structure is good with respect to this value if it *preserves* truth, not if it *produces* truth from falsehoods. Analogously, in asking about the epistemic value of perceptual processes, we are again interested in the value of truth-conducivity (at least broadly understood). And in this theoretical context, where sensory perceptual processes are at issue, we are at least partly concerned with whether the output (experiences) of those processes are driven in some objective way by the stimulus (where the stimulus is some object, event, or property of the world). And to sensibly ask this question, one considers a number of cases involving the same stimulus type and asks whether perceptual processes deliver outputs that are driven (at least largely) by that stimulus. Accordingly, it would seem that for any one instance, to be compared to all of the cases in that range, one needs to assume the desired value of the input which in this context implies that the input must *just be* a stimulus of the same type. If you change the stimulus, as Hohwy is suggesting, then one can no longer ask about the relevant value of the process, namely, whether it outputs experiences that are appropriately sensitive *to that* stimulus type. See Vance 2015b for more on cognitive penetration and the appropriate role of the stimulus in perception. See Goldman 2008, Lyons 2011, and Vance unpublished MS-b, for more on reliabilism and the epistemology of cognitive penetration.
[13] Below, we offer further discussion of what makes a possible case of cognitive penetration "interesting".

framework. Hohwy writes that "[t]his premise rests rather implausibly on equating bottom-up sensory input with prediction error, leaving out a role for uncertainty and noise. Under relatively noisy and uncertain conditions, the perceptual input may underdetermine perceptual inference and give rise to situations where expectations make perceptions differ without much systematic suppression of prediction error" (2013: 123).

First we distinguish noise from uncertainty. Noise involves irregularities and failures of correspondence between the sensory signal and target properties in the world. Noise arises from at least two sources. Externally, it arises from poor environmental conditions, as when you're viewing objects in fog. Internally, noise arises from variability in neural activation, as when neurons fire differently on two occasions, despite the same relevant initial conditions.

Noise in each of the above senses can be further distinguished from uncertainty in perceptual processing. Here again there are at least two kinds of uncertainty of interest. The first type of uncertainty concerns expectations of precision in the sensory signal. These expectations reflect the system's confidence in the sensory signal as a reliable guide to target properties in the world. When this kind of uncertainty is high, the system expects the signal to be imprecise. The second type of uncertainty concerns the distribution of prior hypotheses. In the message passing process that characterizes PCF, each layer of the system has a discrete set of best guesses approximating a prior distribution of relevant hypotheses at that layer. To the extent that the confidence in the various hypotheses at each layer is low, the system exhibits a second kind of uncertainty.

Grant that Hohwy's reply is apt: Premise 2 does conflate two distinct things—prediction error and sensory input. And it is certainly true that the PCF marks a clear distinction here. What's more instructive is Hohwy's appeal to perceptual contexts involving noise, ambiguity, and uncertainty. This appeal is interesting in part because it is not Hohwy's alone. For example, Gary

Lupyan makes a similar proposal within a PCF framework; he writes, "When the low-level information is more ambiguous, high-level knowledge can dominate phenomenology [via cognitive penetration of perception]" (Lupyan 2015: Sect. 5.1)

From within the predictive coding framework, there are at least two reasons to think that noise and uncertainty would tend to lead to CP. The first reason pertains to the conditions under which higher levels of the hierarchy are engaged. In PCF, the bottom-up sensory signal is constituted by an error signal carrying information about the mismatch between actual lower level activity and predicted activity at that level. In cases where the sensory input from the world at the receptors is noisy, the system will tend to predict the signal with less accuracy. Less accurate predictions lead to a larger error signal sent up the hierarchy. The error signal continues to propagate higher if and only if the system continues its inaccurate predictions at the levels to which the signal is sent. If the system accurately explains away the signal at lower levels, there is no error signal to propagate to higher cognitive levels; the error signal becomes too attenuated to engage cognitive priors and cognitive penetration does not occur. But if the error signal persists even to high levels of the hierarchy, cognitive priors are engaged, and cognitive penetration can occur.

The second reason to expect noise and uncertainty to lead to CP pertains to making room in the message passing process for cognitive states to have an influence at lower levels. In the message passing process that characterizes PCF, the system starts with a discrete set of best guesses approximating a prior distribution of relevant hypotheses at each level. Then the system communicates between the levels to help constrain these best guesses and yield a determinate hypothesis at each level for representation in perceptual experience. On this picture, there may be limitations for how higher-level states can influence low-level perceptual representation. In particular, it's possible that cognitive states can influence low-level representations of perceptible properties only if the hypothesis space remains open with respect to those low level properties and the hypothesis favored at the higher levels remains a contender. That is, perhaps cognition cannot override a perceptual inference that is settled at some lower level, but cognitive states could help the system settle on some content at low levels if the hypothesis remains suitably open (Vance unpublished MS-a).

Hohwy utilizes both of the above points in his explanation for why cognitive penetration sometimes does and sometimes does not occur in PCF. He writes,

> The proposal is then that cognitive impenetrability occurs when prediction error is sufficiently suppressed at relatively early [low level] processing stages. Even if the higher-level belief is part of a model generating predictions, these predictions generate little prediction error because activity at lower levels is already very low...One prediction flowing from this proposal is then that if more noise were introduced into the situation, for example by degrading the stimuli or providing a context suggesting added ambiguity, then the prior high-level beliefs...could be allocated more weight [and could lead to cognitive penetration]. (Hohwy 2013: 126).

The above points relate to noise and uncertainty as follows. The system's hypotheses at low levels are more likely to remain open the more uncertain the system is about representations at those low levels. For example, suppose that a subject views an image of a face in shadow (Lee & Mumford 2003). During initial processing at low levels, the system may be uncertain how to represent various differences in luminance across the scene. In particular, at low levels the system may be uncertain whether to represent some boundary line as a shadow or as the edge of an object. According to PCF, the hypothesis space for how to represent the relevant features can remain open as the sensory signal is sent upward to engage higher-level priors. Suppose that, once the high level priors have been engaged, the clear best guess is that the stimulus is a face and the boundary in question is the edge of the jaw. In that case, the higher-level face-relevant perceptual representations will send face-consistent predictions to lower levels and help them converge on a representation that is consistent with the face hypothesis favored at the high level. For example, the higher levels may help constrain the lower levels to converge on a representation of the boundary as an edge of the face rather than a mere shadow edge. If the system had already settled on the low-level representation of the boundary as a shadow, it may have been impossible for the higher level states to influence the lower level by helping them converge on the edge representation. Thus, the more the lower level hypothesis spaces can remain open to influence (due to noise or uncertainty), the more room there is for cognitive penetration of representations at these lower levels.

Hohwy thus suggests that "interesting" cases of cognitive penetration "tend to occur when input is rather uncertain, noisy, or ambiguous" (2013: 137). And although he does not claim that noise and the like are necessary conditions for cognitive penetration (within the PCF), his various comments suggest that he does commit to a claim that *most* cases of CP, or most *interesting* cases of CP, occur only when there is relatively high noise in the incoming signal. He writes, "The view of cognitive penetrability I have advocated predicts that we will find *interesting* cases of penetrability when the perceptual situation is rather riddled with uncertainty: when we are still learning, or when there is noise or ambiguity" (2013: 129; emphasis added). He suggests that "this is borne out by the types of case often used as examples of cognitive penetrability: these cases often involve noisy or otherwise unexpected sensory inputs, or sensory inputs that have not been properly learned yet, or illusions, most of which are based on inherently ambiguous stimuli" (2013: 124).[14] Our concern, to be developed just below, is this. Hohwy is correct to identify *some* interesting putative cases of cognitive penetration with contexts of noise or uncertainty, and indeed the PCF seems to most naturally accommodate and explain some of *these* cases, and in the two ways we clarified just above. However, if it is quantified with any strength—for example, such that (interesting) instances of CP

[14]Hohwy's references to the noisiness of the "sensory inputs" and to "inherently ambiguous stimuli" suggest that he has in mind something close to what we below call the Noise thesis, since the features leading to cognitive penetration are supposed to be inherent in the stimulus itself. However, Hohwy moves between this emphasis and references to uncertainty, which is a feature of the processing system, not the stimulus. As a result, we think it most charitable to allow that Hohwy's explanation appeals both to features of the stimulus and to how the information is processed by the perceptual system. Since these loci of explanation are importantly different, we distinguish them in our discussion and address them both.

tend to occur in noisy and uncertain contexts, or that *most* (interesting) cases of CP are in contexts involving noise and uncertainty—then we suggest that the claims fail to adequately account for cognitive penetration in PCF.

Consider once more one of Hohwy's own descriptions of his view: "interesting" cases of cognitive penetration "tend to occur when input is rather uncertain, noisy, or ambiguous" (2013: 137. On one reasonable disambiguation, the claim is universally quantified: all interesting cases of CP occur in a perceptual context involving substantial noise or uncertainty. But the claim is problematic even if understood more weakly. So let us suppose that the claim is this: most interesting cases of cognitive penetration occur in a perceptual context involving substantial noise or uncertainty. Here 'noise' covers both external and internal noise as described above. 'Uncertainty' covers both expected imprecision of the sensory signal and distribution of confidence levels over the hypotheses about the environment. Since we have distinguished noise from uncertainty, it will be helpful to formulate two distinct theses about cognitive penetration in PCF, one for noise and one for uncertainty, as follows:

> **Noise thesis**: most interesting cases of cognitive penetration occur because noise in the perceptual context is at a substantially larger amount than usual.
>
> **Uncertainty thesis**: most interesting cases of cognitive penetration occur because uncertainty within the subject's perceptual system is at a substantially larger amount than usual.

The use of 'substantial' indicates that the noise or uncertainty involved comes in substantially higher amounts than similar stimuli where cognitive penetration does not occur (see below for more on the relevant contrast cases). Bayesian approaches to perceptual processing (including PCF) are often motivated by appeal to allegedly ubiquitous noise and uncertainty in perceptual processing; the Bayesian inferences help resolve an underdetermination problem arising from such noise and uncertainty.[15] The Noise and Uncertainty theses have to specify an unusually high amount of noise or uncertainty because they are part of an explanation for why cognitive penetration *sometimes* but not always occurs. If the theses pertained to ubiquitous noise and uncertainty, respectively, they would be inapt for explaining why cognitive penetration sometimes does *not* occur.

In what follows, we argue that the Noise thesis is false.[16] Concerning the Uncertainty thesis, we argue that it may be true; however, even if it is true, it fails to explain why cognitive penetration sometimes does and sometimes does not occur in PCF.

We turn first to the Noise thesis. Here are two large sets of possible cases of cognitive penetration, involving a broad range of stimulus types, that suggest that the Noise thesis is false. The first is largely empirical, the second, conceptual or philosophical.

As many readers of this journal will know, there has been a resurgence in the last decade or so in philosophical discussion of the possible cognitive penetration of perception. What's important about these analyses for present purposes is that they move beyond merely hypothetical or anecdotal cases, and provide instead rather detailed interpretations of experiments and data from perceptual psychology. Here are a few such experimental cases that have received recent philosophical attention, and typically from a number of theorists. In each case, the feature to note is the relevant experimental stimulus or stimuli. To start, early research by New Look psychologists continues to receive attention. In perhaps the most famous example, the experimental condition involved extended exposure to regular coins of varying values, where subjects matched an adjustable light patch to match the stimulus in real time. A recent experiment explicitly followed this experimental paradigm, where subjects performed a real-time size-matching task using computer images of

---

[15] Feldman 2015, Kersten and Yuille 2003; Rescorla 2013, Vance 2015b.
[16] See Vance 2015a for another argument against the Noise thesis, though not under that name.

objects that typically correlate with a negative, positive, or neutral emotional valence.[17] Another set of experiments involved control stimuli of paper cutouts of familiar objects associated with a particular color—for example a love-heart shape—which were all cut from a uniformly orange piece of paper. Here subjects were asked to color-match the cutout images (where controls were images with no characteristic color) to a colored background.[18] In more recent work, subjects were asked to adjust computer images of objects with characteristic colors—a yellow banana, a blue Smurf, a red Coca-Cola icon—to a subject-relative "perfect grey".[19] Another approach has been to focus on empirical research concerning eye-tracking and visual spatial constancy. In these studies, the experimental stimuli range from simple moving dots to realistic paintings.[20] Finally, in a recent set of studies, experimental subjects were asked to match the luminance, in real time, of greyscale images of faces stereotypical of black males or white males. And in a second iteration of the study, subjects were asked to perform roughly the same task with an image of a racially ambiguous[21] male face, and when coupled with a 'WHITE' or 'BLACK' linguistic cue.[22] Now, all of these cases are controversial as genuine cases of cognitive penetration, even if they are some of the cases that have received the most thorough and compelling CP interpretations. But what's relevant here is that in none of these studies are the experimental (or control) stimuli noisy, ambiguous, uncertain, or messy in any way beyond those of stimuli that do not elicit putative cognitive penetration. In each case, the stimuli are instead clearly visible, and presented "online" synchronously with the report mechanism (for example, an adjustable computer display). And while these cases do not exhaust those that have been alleged as instances of cognitive penetrability, they do make up a substantial portion of those

---

[17] Bruner and Goodman 1947, van Ulzen et al. 2007. See Stokes 2012, 2013 for discussion.
[18] Delk and Filenbaum 1965. See Macpherson 2012; Stokes 2013 for discussion.
[19] Hansen et al. 2006, Olkkonen et. al 2008; Witzel et al. 2011; See Stokes 2014 and Stokes and Bergeron 2015 for discussion.
[20] Chelazzi et al. 2001; Yarbus 1967; Land 2006; Wu, forthcoming.
[21] Note that this stimulus set was racially ambiguous, but not ambiguous in any of the relevant senses used by Hohwy. That is, the image was not noisy, or occluded, or difficult to make out in any sense.
[22] Levin & Banaji 2006. For philosophical discussion, see Macpherson 2012; Stokes and Bergeron 2015.

cases that have received sustained analysis and debate. The problem for Hohwy's view should be clear: here we have several putative cases of cognitive penetration (indeed, one might say several *types* of experimental case) but none of which involve substantial noise. As an empirical matter, the Noise thesis looks dubious.

The same worry can be expressed on more conceptual grounds. In the middle of the 20th century, N.R. Hanson and Thomas Kuhn challenged the empiricist assumption that perceptual observation (as well as report) in science was theoretically neutral. The concern was that scientists with robust theoretical commitments—one might here think of Kuhn's "normal scientists"—may visually or otherwise perceptually experience the world, whether in the wild or in the lab, in ways substantially biased by those background theoretical commitments. To choose one amusing, even if merely hypothetical case, the theorist of heliocentrism might visually observe the setting sun in ways different from, even though standing side by side with, the theorist of geocentrism. This was supposed to be an important general challenge because it threatened the epistemic role of perceptual observation. Put simply, how can perceptual observation support rational theory choice, and adjudicate scientific dispute, if it is laden with theoretical beliefs and commitments? Most of the cases invoked by these philosophers were imaginary or hypothetical. However, many of the examples were grounded in actual scientific practice (for instance, appealing to actual technologies and techniques of scientific observation), and others made explicit appeal to research on perception by the contemporary New Look psychologists.

Here again, whether perceptual observation *is* theory-laden in these ways, and in the kinds of cases adduced by Hanson and Kuhn, remains controversial. But, again, for our purposes the crucial feature to note is that few if any of the candidate cases involved substantial noise (or uncertainty). In fact, quite the contrary: these are cases where highly trained individuals are performing the tasks standard for their domain of expertise, but where that very expertise may be biasing or otherwise affecting perceptual experience in that domain. (And the phenomenon would appear to generalize; one might think of the perceptual expertise that comes in any domain involving distinctive technology, for example, medical imaging techniques, auto mechanics, mass media production.) These are perceptual situations not riddled with noise, but often very nearly devoid of it. So, here again we have a wide range of cases that would seem to be missed by the Noise thesis.

In the examples of putative cognitive penetration above, the amount of external and internal noise is not greater than in similar cases of non-cognitive penetration found in control cases from the same studies. For example, in the memory color studies by Hansen and colleagues, in the experimental condition, subjects viewed images with familiar, characteristic color (e.g. bananas, Coke logos, Smurfs). In control conditions, subjects viewed images without a characteristic color (e.g. socks, golf balls, mushrooms,). Both sets of stimuli were clearly displayed on computer screens in good lighting conditions. External noise was thus at similar amounts for each kind of stimulus and was not greater than it would be for normal viewing conditions. In addition, since experimenters in studies such as these ensure that all participants enjoy normal visual acuity, there is no reason to think that there was an unusually large amount of internal noise in the putative cognitive penetration cases as compared to control cases in the same studies.

We now turn to the Uncertainty thesis. Regarding uncertainty, the PCF theorist could hold that, in the cases in question, subjects' perceptual systems had greater than normal amounts of uncertainty in low levels of processing. In the memory color studies by Hansen et al. the methodology required subjects to adjust the images to subjective perfect grey. Stimuli in these cases have several relevant properties, including color and shape. Some of the objects with a familiar shape had a characteristic color far from grey (e.g. Coke logos are bright red; banana-shaped objects are yellow). As subjects adjusted the image to make it greyer and greyer, subjects' perceptual systems may have expected that the sensory signal was getting less and less precise. So it is reasonable that by the time the computer screen displayed a grey Coke logo, say, the perceptual system took the sensory signal to be very imprecise. This could have led to uncertainty either in the sense of expected imprecision in the signal or in the distribution of confidence-levels over relevant low level hypotheses (e.g. concerning the color of the image). For example, the high-level expectation of an imprecise sensory signal could have had a downward influence, leading to lower-level expectations of an imprecise signal. In addition, this could have caused lower-level hypotheses concerning the color of the object to be less certain. Similar points could be made for the ambiguous race faces. Since there was ambiguity about which race the face belonged to, the theorist of PCF could say that subjects' perceptual systems were less certain at low levels about whether the face had the luminance it did because the person had darker skin or because of lighting conditions. The above points support the claim that uncertainty was involved in making cognitive penetration in such cases possible, because uncertainty at low levels helped leave room for influence by cognitive priors.

However, even if we grant the top-down influence by high-level priors on lower-level processing resulted in low-level uncertainty in the system, the above line of response leaves a gap in the PCF explanation for why cognitive penetration sometimes occurs and at other times does not. Suppose that subjects' perceptual systems were uncertain in their processing of the low level properties of the objects in question (e.g. the color of the Coke logo). On the proposal we sketched above, the content of the prior in question links a shape (possessed by Coke logos) with a color (bright red). Call this the prior that Coke logos are bright red, though note that the content need not involve the concept COKE LOGO, for it could include only the relevant shapes and colors. The needed low-level uncertainty could then be caused by the engagement of a prior linking the shape of the stimulus to its color. If the uncertainty arose from a prior that Coke logos are bright red and the expectation that the greyness of the stimulus was due to poor viewing conditions, that would have required the system to engage the high-level prior about the stimulus being a Coke logo before it became uncertain concerning the stimulus conditions. However, in Hohwy's use of the Uncertainty thesis to explain cognitive penetration, things go the other way around: uncertainty at low levels of processing occurs first and explains why the high-level priors are engaged. Recall that on Hohwy's explanation, lower level uncertainty about the stimulus causes an increased error signal to propagate up the hierarchy and engage cognitive priors, which then help resolve the uncertainty through cognitive penetration (Hohwy 2013: 126). So, on Hohwy's explanation, the relevant high-level priors are engaged *after* the relevant uncertainty occurs and the occurrence of the uncertainty explains why the priors are engaged. By contrast, the explanation we sketched above requires that the relevant priors are engaged *before* the relevant uncertainty occurs and the uncertainty does not explain why the priors are engaged. Hohwy's explanation thus leaves unexplained why the relevant priors are engaged in the cases of CP we discuss. So even if Hohwy could say that there are greater than normal amounts of uncertainty in perceptual processing concerning the stimulus conditions and color of the objects in the Hansen et al. studies, this won't help his appeal to uncertainty to explain why cognitive penetration sometimes does and at other times does not occur, according to PCF.

Furthermore, it will not save Hohwy's analysis to appeal to the use of 'interesting' in the Noise and Uncertainty theses. As noted above, the notion of 'being interesting' at play is not sufficiently motivated. It is at one point suggested that the principle for identifying the interesting cases of cognitive penetration is given by appeal to a kind of rational need to draw perceptual inference in circumstances where the sensory input underdetermines that inference. Cognitive priors then exert top-down influence and satisfy that rational need "to do something" (inferentially), filling in what sensory input has left open. What circumstances are these? Well, just those involving substantial noise or uncertainty (see Hohwy 2013: 123-4). Clearly this won't do as a defence against a worry that the analysis, with its commitment to the Noise and Uncertainty theses, fails to explain many types of case. That worry said that there is a wide range of plausible cases of the phenomenon that do not involve any substantial noise or uncertainty. To reply that the analysis on offer is only intended to explain "interesting" cases, and then define the value of interest in terms of noisy perceptual situations, would be nothing short of circular.

The more plausible, and broadly motivated, way to understand interest-value is to appeal to possible consequences of candidate cases of cognitive penetration. This was discussed above, so a brief reminder suffices here. Traditionally, going as far back as the philosophers of theory-ladenness already discussed, one such consequence is epistemic. Possible cases of cognitive penetration are supposed, by friends and sceptics of CP alike, to be important for how we theorize the knowledge-providing role of perception, in both scientific and everyday empirical contexts. More recently, a possible consequence of importance is scientific. If perception is cognitively penetrable, and perhaps relatively pervasively, then modular theories of the architecture of perception may require substantial revision. All of the candidate cases above—both experimental and conceptual/anecdotal—would plausibly bear one or both consequences of interest. And none of them appear to involve substantial noise or uncertainty (in the senses we have clarified). So on an extant characterization of the relevant interest-value, Hohwy's analysis fails (or at least looks challenged, given commitment to the Noise and Uncertainty theses) to explain a wide range of *interesting* cases.

V. Conclusion

As we have noted, the putative examples of cognitive penetration described in the previous section are controversial. Many theorists deny that these are in fact cases of cognitive penetration, properly defined. In addition, following Fodor, Pylyshyn and others, many theorists take the dialectical burden of proof to be on the shoulders of those who claim that cognitive penetration occurs. They take the default assumption to be that perceptual systems are largely modular in Fodor's (1983) sense. The standard strategy in arguing for cognitive penetration is to propose specific examples for consideration and to argue that it is difficult to explain the phenomenon without acknowledging some interesting cognitive effect on perception: cognitive penetration.

However, the development of non-modular mental architectures, such as that proposed in PCF, raises a possible burden shift in the dialectic. If modularity is the default assumption, then one needs clear and convincing positive evidence for cognitive penetrability on a case by case basis. But if a non-modular architecture in the form of a predictive coding framework were the default assumption, one would not need to argue for cognitive penetration in the same way, since the now-default architecture would suggest the existence of cognitive penetration independently of the details of any particular example. We think that this would be interesting theoretical progress, and progress that would potentially break from decades of rather entrenched cognitive scientific theorizing about cognitive effects on perception. In other words, this is an interesting and relatively untapped consequence of the bold new view offered by predictive coding. We would only hasten to add that, as our critical analysis above hopefully indicates, the PCF would need significantly more development and support than it currently enjoys, in order to bear this kind of dialectical fruit.

REFERENCES

Ashby, W. R. (1947). Principles of the self-organizing dynamic system. *The Journal of General Psychology*, *37*(2), 125–128. http://doi.org/10.1080/00221309.1947.9918144

Bruner, J. S., & Goodman, C. C. (1947). Value and Need as Organizing Factors in Perception. *Journal of Abnormal and Social Psychology*, *42*(1), 33–44.

Chelazzi, L., Miller, E., Duncan, J., Desimone, R. (2001)"Responses of Neurons in Macaque Area V4 during Memory-Guided Visual Search," *Cerebral Cortex* 11(8): 761–72

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science . *Behavioral and Brain Sciences*, *36*(03), 181–253. http://doi.org/10.1017/S0140525X12000477

Clark, A. (2015). *Surfing uncertainty: Prediction, action, and the embodied mind*. Oxford University Press.

Cole, S., Balcetis, E., and Dunning, D. (2013) Affective signals of threat increase perceived proximity. *Psychological Science*, 24(1), 34-40.

Danks, D. (2014). *Unifying the mind: Cognitive representations as graphical models*. MIT Press.

Delk, J. L., & Fillenbaum, S. (1965). Differences in Perceived Color as a Function of Characteristic Color. *The American Journal of Psychology*, *78*(2), 290–293.

Egner, T., Monti, J. M., & Summerfield, C. (2010). Expectation and surprise determine neural population responses in the ventral visual stream. *The Journal of Neuroscience*, *30*(49), 16601-16608.Friston, K. (2005). A theory of cortical responses. *Philosophical transactions of the Royal Society B: Biological sciences*, *360*(1456), 815-836.

Feldman, J. (2015). Bayesian models of perception: a tutorial introduction. In J Wagemans (Ed.) *The Oxford handbook of perceptual organization.* Oxford University Press.

Fodor, J. A. (1983). The Modularity of Mind. MIT Press.

Fodor, J. (1984). Observation reconsidered. *Philosophy of science*, 23-43.

Fodor, J. A. (1988). A Reply to Churchland's' Perceptual Plasticity and Theoretical Neutrality'. *Philosophy of Science*, *55*(2), 188–198. http://doi.org/10.1086/289426

Friston, K. (2009). The free-energy principle: a rough guide to the brain?. *Trends in cognitive sciences*, *13*(7), 293-301.

Hohwy, J. (2013). *The predictive mind*. Oxford University Press.

Goldman, A. (2008). Immediate justification and process reliabilism. In Q. Smith (Ed.) *Epistemology: New essays*, (63-82). Oxford University Press.

Hansen, T., Olkkonen, M., Walter, S., & Gegenfurtner, K. R. (2006). Memory modulates color appearance. *Nature Neuroscience*, *9*(11), 1367–1368. http://doi.org/10.1038/nn1794

Helmholtz, H. v. (1867). *Handbuch der physiologishen optic*. Leipzig: Leopold Voss.

Hinton, G. E., Dayan, P., Frey, B. J., & Neal, R. M. (1995). The" wake-sleep" algorithm for unsupervised neural networks. *Science*, *268*(5214), 1158-1161.

Hohwy, J. (2013). The Predictive Mind. Oxford: Oxford University Press. http://doi.org/10.1093/acprof:oso/9780199682737.001.0001

Kersten, D., & Yuille, A. (2003). Bayesian models of object perception. *Current opinion in neurobiology*, *13*(2), 150-158.

Land, M.F. (2006) "Eye Movements and the Control of Actions in Everyday Life," *Progress in Retinal and Eye Research* 25: 296–324

Lee, T. S., & Mumford, D. (2003). Hierarchical Bayesian inference in the visual cortex. *JOSA A*, *20*(7), 1434-1448.

Levin, D. T., & Banaji, M. R. (2006). Distortions in the perceived lightness of faces: The role of race categories. *Journal of Experimental Psychology: General*, *135*(4), 501–512. http://doi.org/10.1037/0096-3445.135.4.501

Lupyan, G. (2015). Cognitive penetrability of perception in the age of prediction: Predictive systems are penetrable systems. *Review of Philosophy and Psychology.*

Lyons, J. (2011). Circularity, reliability, and the cognitive penetrability of perception. *Philosophical Issues*, *21*(1), 289-311.

Macpherson, F. (2012). Cognitive Penetration of Colour Experience: Rethinking the Issue in Light of an Indirect Mechanism. *Philosophy and Phenomenological Research*, *84*(1), 24–62. http://doi.org/10.1111/j.1933-1592.2010.00481.x

Mumford, D. (1992). On the computational architecture of the neocortex. *Biological cybernetics*, *66*(3), 241-251.

Olkkonen, M., Hansen, T., & Gegenfurtner, K. R. (2008). Color appearance of familiar objects: Effects of object shape, texture, and illumination changes. *Journal of Vision*, *8*(5), 13–13. http://doi.org/10.1167/8.5.13

Pylyshyn, Z. (1999). Is vision continuous with cognition? The case for cognitive impenetrability of visual perception. *Behavioral and Brain Sciences*, *22*, 341–423.
Raftopoulos, A. and Zeimbekis, J. (2015) "The cognitive penetrability of perception: An overview" In J. Zeimbekis & A. Raftopoulos (Eds.), *The Cognitive Penetrability of Perception* (pp. 75–100). Oxford: Oxford University Press. http://doi.org/10.1093/acprof:oso/9780198738916.003.0003
Rescorla, M. (2013). Bayesian perceptual psychology. *Oxford Handbook of the Philosophy of Perception*. Oxford University Press.
Siegel, S. (2012). Cognitive penetrability and perceptual justification*. *Noûs*, *46*(2), 201-222.
Silins, N. (2016). Cognitive penetration and the epistemology of perception. *Philosophy Compass, 11*(1), 24-42.
Stefanucci, J.K. and Proffitt, D.R. (2009). The roles of altitude and fear in perception of heights. *Journal of Experimental Psychology: Human Perception & Performance*, 35, 424-438.
Stefanucci, J.K. (2010) Emotional high: Emotion and the perception of spatial layout. In E. Balcetis and G.D. Lassiter (Eds.), *The Social Psychology of Visual Perception*, Psychology Press (pp.1890-1902).
Stefanucci, J.K., Gagnon, K.T., & Lessard, D. (2011) Follow your heart: Emotion adaptively influences perception. *Social Psychology and Personality Compass*, 5/6, 296-308.
Stefanucci, J.K., Gagnon, K.T., Tompkins, C., Bullock, K.E. (2012). Plunging into the pool of death: Imagining a dangerous outcome influences distance perception. *Perception*, 41, 1-11.
Stokes, D. (2012). Perceiving and desiring: a new look at the cognitive penetrability of experience. *Philosophical Studies*, *158*(3), 477–492. http://doi.org/10.1007/s11098-010-9688-8
Stokes, D. (2013). Cognitive Penetrability of Perception. *Philosophy Compass*, *8*(7), 646–663. http://doi.org/10.1111/phc3.12043
Stokes, D. (2014). Cognitive Penetration and the Perception of Art. *Dialectica*, *68*(1), 1–34. http://doi.org/10.1111/1746-8361.12049
Stokes, D. (unpublished MS). Attention and the cognitive penetration of perception.
Stokes, D. (2015). Towards a consequentialist understanding of cognitive penetration. In J. Zeimbekis & A. Raftopoulos (Eds.), *The Cognitive Penetrability of Perception* (pp. 75–100). Oxford: Oxford University Press. http://doi.org/10.1093/acprof:oso/9780198738916.003.0003
Stokes, D., & Bergeron, V. (2015). Modular architectures and informational encapsulation: a dilemma. *European Journal for Philosophy of Science*, *5*(3), 315–338. http://doi.org/10.1007/s13194-015-0107-z
van Ulzen, N. R., Semin, G. R., Oudejans, R. R. D., & Beek, P. J. (2007). Affective stimulus properties influence size perception and the Ebbinghaus illusion. *Psychological Research*, *72*(3), 304–310. http://doi.org/10.1007/s00426-007-0114-6
Vance, J. (2014). Emotion and the new epistemic challenge from cognitive penetrability. *Philosophical Studies*, 169(2), 257-283.
Vance, J. (2015a). Review of *The predictive mind* by Jakob Hohwy. *Notre Dame Philosophical Reviews*
Vance, J. (2015b). Cognitive penetration and the tribunal of experience. *Review of Philosophy and Psychology*, *6*(4), 641-663.
Vance, J. (in preparation). Prediction, action, and revision. In *MIND23*. T. Metzinger and W. Wiese (Eds.)
Vance, J. (unpublished MS-a). Illusion and the predictive mind.
Vance, J. (unpublished MS-b). Mentalism, reliabilism, and cognitive penetration.
Vetter, P., & Newen, A. (2014). Varieties of cognitive penetration in visual perception. *Consciousness and cognition*, *27*, 62-75.

Witzel, C., Valkova, H., Hansen, T., & Gegenfurtner, K. (2012). Object knowledge modulates colour appearance. *I-Perception*, *2*(1), 13–49. http://doi.org/10.1068/i0396

Wu, W. (forthcoming). Shaking up the mind's ground floor: The cognitive penetration of visual attention. *Journal of Philosophy*

Yarbus, A.L. (1967) *Eye Movements and Vision*, New York: Plenum Press

Yedidia, J. S., Freeman, W. T., & Weiss, Y. (2003). Understanding belief propagation and its generalizations. *Exploring artificial intelligence in the new millennium*, *8*, 236-239.