

Perceiving as Knowing in the Predictive Mind

Daniel Munro

Penultimate draft; please cite published version.

1. Introduction

Aside from the occasional external world skeptic, almost every philosopher agrees that perception is a source of knowledge about the world, with many taking it to be the most foundational source. But a more contentious issue concerns exactly what role knowledge should play in how we understand the very nature of perception.

On one popular picture, knowledge has little relevance for understanding the intrinsic nature of perceptual experience *qua* mental state. This ‘internalist’ picture understands all perceptual experiences as a matter of how things subjectively seem to one or how one takes things to be, which can be characterized independently of whether perception is tracking the truth and yielding knowledge. While internalists agree that perception is typically a source of knowledge, they hold that knowledge is not itself a mental state but a hybrid of the mental state of belief plus other, non-mental factors (truth, justification, etc.). On this picture, a subject whose perceptual experience yields knowledge can be mentally identical to one whose doesn’t (e.g., one who is hallucinating) if everything experientially *seems* identical to each. Knowledge thus isn’t relevant to understanding the intrinsic nature of perception since we can give a complete characterization of any given perceptual state without mentioning knowledge.

There’s an alternative picture, though, on which perception can’t be fully understood independently of knowledge. The ‘knowledge first’ picture holds that some mental states are constituted by relations between a subject and her environment, rather than all mental states being ‘internal’ to a subject’s mind. Knowledge is the paradigmatic such state. Furthermore, successful perceiving is itself a *way of knowing*, or a knowledge-constituting relation between subject and environment. Since successful perceiving is one particular way of occupying the mental state of knowledge, two subjects for whom everything experientially *appears* the same may nevertheless not be mentally identical. Furthermore, any experiential state which doesn’t constitute knowledge is conceptualized as a kind of botched attempt by one’s cognitive system to yield knowledge. Knowledge is therefore part of any complete account of the nature of perception itself, unlike the internalist picture on which knowledge is merely a product which perceptual experiences can yield.

In opposition to the knowledge first picture, philosophers sometimes claim that knowledge has no place in a naturalistic or scientifically respectable picture of the mind, including in how we

understand mental states like perceiving. Such philosophers argue that the intrinsic nature of one's mental states doesn't depend on how a subject is related to the world, since we could vary the states of the world around a subject while how things seem to her and the neurocognitive processes 'inside her head' remain exactly the same (see, e.g., Fodor 1980; Churchland 1987; Molyneux 2007; Pernu 2009; Ganeri 2010; Michaelian 2016).¹ If this line of argument is correct, it suggests that we should accept the internalist picture of perception over the knowledge first one.

This paper considers these issues in light of one of the most powerful recent frameworks for scientific study of the mind: the 'predictive processing' (PP) framework. PP sees the brain's fundamental organizing principle as *prediction error minimization*, in which the brain generates predictions about the states of the world and revises them in the face of errors. According to PP, all of perception, action, and cognition emerge from such processes. Perception, specifically, is a process in which the brain makes predictions about the world around itself and uses stimulations to sensory organs—for example, retinal stimulation in the case of vision—as feedback to revise these predictions when they're in error. Perceptual experience occurs when the brain settles on the set of hypotheses that best minimizes prediction error. In this picture, then, experience is a matter of the brain actively constructing a representation of the world, and the role of sensory stimulation is to reveal when this representation is in error and should be revised. This kind of computational framework has seen impressive success in explaining and unifying various mental phenomena² and has been increasingly applied to explain physical neuroanatomy (Friston et al. 2014; Parr and Friston 2018).

PP's proponents sometimes describe the framework in a way that suggests we should oust knowledge from our theorizing about perception. Because PP says that perception involves the brain constructing an internally-generated representation, its proponents have described perceptual experience as a 'fantasy that coincides with reality' (Frith 2007, 135), an 'inferred fantasy' (Paton et al. 2013, 222), a 'virtual reality' (Hohwy 2013, 137; Hobson and Friston 2014), and as being 'secluded

¹ Philosophers of this tendency often trace their inspiration to Quine (1969). A related line of argument, also often traced to Quine, focuses not on truth but on normativity: knowledge is a normative notion because it entails notions like justification; since sciences of the mind deal in descriptive facts, normative facts are somehow non-naturalistic. This paper won't directly address this line of thought; for resistance to it, see, e.g., Kim (1988) and Jenkins (2007).

² For example: perceptual attention (Clark 2016, ch. 2); various perceptual phenomena and illusions (Hohwy 2013, ch. 5); the nature of consciousness (Clark, Friston, and Wilkinson 2019); hallucinations and delusions in conditions like schizophrenia (Benrimoh et al. 2018); the imagination (Kirchhoff 2018); emotions (Seth 2013); and various elements of social cognition, including mindreading in autism (Palmer et al. 2015) and stereotyping (Westra 2019).

from the world' (Hohwy 2016, 259). These descriptions may seem to suggest that PP makes perception a process that puts us in touch with representations constructed by our brains rather than with mind-independent external reality. If that's right, perhaps we don't need to bring knowledge into the picture in order to understand the nature of perception. Rather, perceptual experience is a state that's internal to the mind, and any alignment between what we experience and the world is a further, non-mental fact.

At the same time, there are also features of PP which on closer inspection seem aligned with the knowledge first picture. For one thing, perceptual processing under PP is sometimes described as, at its core, a process by which one's mind is 'entrained by' one's environment, with the dynamics of the brain's unfolding predictions recapitulating the dynamics of one's environment. This sounds like one possible way of cashing out the idea that knowledge-constituting perception is fundamentally a factive relation between mind and environment. Furthermore, Clark (2016, ch. 3) argues—against the oft-repeated slogan that PP makes perception a kind of 'controlled hallucination'—that PP makes hallucination a kind of *uncontrolled perception*, since it involves the brain's predictions failing to be coupled in the right way to the world. This seems reminiscent of the knowledge first idea that hallucination is a kind of botched attempt at achieving a knowledge-constituting relation to one's environment, in contrast with the internalist idea that hallucination and successful perception are the same kind of internal mental state.

So, despite initial appearances to the contrary, there seems to be some promise in the idea that PP fits well with the knowledge first picture. This paper makes good on that promise. I argue that, far from implying that we should banish knowledge from our theorizing about the mind, accepting PP's picture of perceptual psychology gives us reason to accept the knowledge first understanding of perception. In other words, if PP is correct, we have reason to accept that the nature of perception should be conceptualized in part in terms of knowledge. It's therefore false that scientifically respectable theorizing about the mind should avoid theorizing about knowledge: PP, a popular scientific framework, supports the idea that knowledge should be central to our theorizing about perception.

Here's the paper's plan. §2 covers some groundwork for the paper's main arguments. It first gives a brief overview of the PP framework; it then sketches the main components of the knowledge first picture of perception in contrast with the rival internalist picture; and it closes by clarifying the respective theoretical domains of PP versus the knowledge first and internalist frameworks. The paper's main arguments begin in §3, where I consider the 'good case,' or cases of successful,

knowledge-yielding perception. I argue that PP gives us an account of the good case that supports the knowledge first picture over the internalist picture. §4 then turns to the ‘bad case,’ or cases in which perception goes awry and fails to yield knowledge. I argue that, again, adopting a PP account of the bad case supports the knowledge first picture over the internalist one.

2. Preliminaries: PP, knowledge first, internalism

2.1. Predictive processing

This subsection introduces the PP framework, sketching its aspects that will be especially relevant for my arguments below.³

On a more traditional account of perception, perceptual processing involves receiving sensory stimulation from the environment (in the form of, e.g., retinal stimulation) and then extracting information about the world from it. It’s thus primarily a matter of simply receiving information in a ‘bottom-up’ way. PP says that, instead, perception is driven in a more active, ‘top-down’ way by the brain’s inferences about its surrounding environment. The brain’s ‘generative model’ encodes information about the world and about the patterns of sensory stimulation different states of the world tend to cause. This model constantly generates predictions about the evolving states of the world around oneself and about the kinds of bottom-up sensory stimulation these states of the world will cause. If this model successfully predicts the stimulation one’s sensory organs receive, its predictions about the worldly causes of that sensory stimulation are deemed successful. If not, prediction error signals tell the brain to revise its predictions. We have experiences when the brain has inferred the set of hypotheses which best minimizes prediction error.

According to PP, we’re never conscious, at the personal level, of the sensory stimulation we’re receiving or of anything like a stream of raw ‘sense data’. Rather, sensory stimulation is, along with the brain’s prior information about the world, one of the sub-personal ingredients used to infer what’s going on in the world around oneself. The brain infers hypotheses about the causes of its sensory stimulation, and (assuming it gets things right) we become conscious of the things in the world which those hypotheses are about.

That’s a rough gloss on perceptual processing in PP, but a more precise detail is important: PP models the way the brain stores and processes information *probabilistically*. The generative model consists of probability distributions over hypotheses about states of the world and the way they cause

³ For more thorough, scientific introductions to the framework, see Rao and Ballard (1999) and Friston (2005); for more philosophical introductions, see Hohwy (2013) and Clark (2016).

sensory stimulations. Rather than just inferring a single set of hypotheses about the world, the brain infers a set of probability distributions over competing hypotheses. Our experience reflects the contents of the hypotheses deemed *most likely*, once prediction error has been minimized. This proceeds via (an approximation of—cf. Wiese 2017) Bayesian computational processes (more precise details of the mathematical framework aren't necessary for my purposes, but see Hohwy 2013, ch. 1 for a helpful overview). To make these computations tractable, though, the brain's inferences can't operate over an infinite hypothesis space. Instead, when predicting the states of the world by generating probability distributions, the brain must begin from a more circumscribed space of 'live' hypotheses under consideration. To do this, the brain relies on abstract, deeply entrenched prior expectations about the world, which bias perceptual expectations and are very resistant to revision (cf. Swanson 2016; Kirchhoff 2018). Plausible examples of such priors are the expectation that multiple distinct objects won't be in the same place at the same time; the expectation that objects don't cease to exist just because we're no longer perceiving them; and certain expectations of 'folk physics' about, for example, how objects in motion behave.⁴

Finally, the PP framework also understands *action* in terms of prediction error minimization. For my purposes, 'epistemic actions' are the most important type of action. Epistemic actions are one way the brain reduces uncertainty in its generative model (cf. Friston et al. 2015; Pezzulo and Nolfi 2019). The most intuitive examples concern more abstract, long-term predictions. For example, in the face of a high degree of uncertainty about whether it's going to rain tomorrow, I can pull out my phone and check a weather app. Doing so allows me to better minimize error when it comes to predicting tomorrow's weather. The PP model of epistemic actions also extends to actions which reduce uncertainty in our low-level perceptual predictions. As we go about the world, we don't merely passively receive a static set of sensory input and then infer its causes. Rather, we are 'very active in the way we go about perceiving the world. We explore, check out, test, look closer, feel, and so on, all of which are ways of actively engaging with the world and thereby changing the sensory input we receive' (Hohwy 2013, 77). Since our generative models include expectations about how sensory input will change as we interact with the world, we can further reduce uncertainty in our perceptual predictions by actively 'sampling' sensory signals. If the sensory input changes as expected given the

⁴ For more examples and scientifically-informed discussion of such abstract expectations in perceptual processing (though not specifically in a PP framework), see Jenkin (2020). It's a further, interesting question exactly where such priors come from; plausibly, some are innate while others are learned over time. On this question, see Clark (2013) and Kirchhoff (2018).

current leading hypothesis about its causes, this provides further confirmation for that hypothesis; if it changes in unexpected ways, we either need to sample the world more or revise that hypothesis. This idea has been applied even to low-level, automatic actions such as eye saccades: our rapid, automatic eye movements are a way of taking in more sensory data from the scene one is looking at, thereby (dis)confirming one's predictions (Parr and Friston 2017; Mirza et al. 2018; Parr et al. 2019).

2.2. *Perceptual knowledge: internalism vs. knowledge first*

This subsection sketches the key differences between the knowledge first and rival internalist pictures of perceptual knowledge.

The differences between the two trace back in part to a disagreement about the nature of the mental, so it's important to first understand this prior disagreement. The internalist subscribes to a picture on which a subject's mental states are exhausted by states which are 'internal' to the mind: how a subject takes things to be, regardless of how things really are; how things seem, appear, or feel to a subject; and so on. Whether a subject possesses such states is independent of whether her mental states track truths about the world. Belief is a paradigmatic example of such a state, because whether a subject believes that P depends only on whether she takes P to be true, not on whether P is in fact true. On this picture, knowledge isn't a mental state in its own right, since whether one knows that P depends on the truth of P, a factor which is taken to be 'external' to a subject's mind. In contrast, the knowledge first proponent holds that, in addition to strictly internal states, some mental states are constituted by *relations between* a subject and environment, or states that relate a subject to truths about her environment. Knowing that P is one such mental state, which relates a subject to the fact that P. This makes knowledge a factive mental state: it's essential to the nature of knowledge that it's a propositional attitude one can have only towards truths (cf. Williamson 2000; Nagel 2017).⁵

These distinct understandings of the mental inform disagreements about the nature of perception and perceptual knowledge. I'll cash out these differences in part in terms of how each picture conceptualizes the 'good case' and 'bad case.' The good case is when everything goes well perceptually, resulting in perceptual knowledge. The bad case is any case in which perceptual experience fails to be a source of knowledge—either because one's experience is non-veridical (as in, e.g., a non-veridical hallucination) or because one's experience isn't causally hooked up to the world

⁵ I don't mean to saddle 'internalist' views with *semantic* internalism. Whether the *contents* of mental states depend on external factors is a distinct issue from whether *types* of states do (but see Williamson 2000, ch. 2 on why reasons paralleling those for the former may favour the latter).

in the right way (as in, e.g., a hallucination that happens to be veridical). I'll use 'successful perceiving' and similar locutions to refer to one's perceptual state in the good case.

Consider again the picture on which all genuine mental states are internal states, which informs internalist accounts of perception. On these accounts, perceptual experiences are, like all other mental states, internal states of a subject. We can divide internalist accounts into roughly two kinds.

On the first kind, the role of perceptual experience is to present a subject with a way things seem or appear to be in the world. How to understand this notion of a perceptual seeming or appearance depends on one's account of the contents of perception. It may be that perceptual experience represents only low-level properties (e.g., colours, shapes, textures), or it may also represent high-level properties (e.g., object categories like *cup* and *dog*). Regardless of how the internalist settles such issues, what's more important here is the relationship between perception and belief: on this version of internalism, how things seem in perceptual experience is the basis on which one subsequently forms beliefs about the world. So, perceptual experience is not a matter of the subject taking the world to be some way; instead, it's a matter of the subject being presented with a way things seem, on the basis of which she forms a belief.⁶

A second sort of internalist account takes experiences to *be* a kind of belief, or at least a very belief-like state; on these accounts, perceptual experience is itself a propositional attitude in which a subject takes the world to be some way, rather than doing so only when one bases a belief on an experience.⁷ On this version of internalism, perceptually experiencing a cup on the table does involve the subject taking it to be the case that there's a cup on the table.

On both kinds of internalism, while our perceptual beliefs typically do amount to knowledge, the mental states of perceptual experience and perceptual belief can be characterized independently of the notion of knowledge, in terms of how things seem for a subject and how a subject takes the world to be, respectively. So, a subject successfully perceiving that P and her 'twin' who merely hallucinates that P are occupying the same mental state, as long as everything experientially seems identical to each.

⁶ Views of this sort can be further divided into at least two kinds. Some are more inferentialist, on which one infers beliefs about the external world from facts about how things seem or appear to oneself (e.g., Schiffer 2009; McGrath 2017). Others have it that our perceptual beliefs are non-inferential, but their basis is nevertheless still how things perceptually seem (e.g., Pryor 2000; Huemer 2006). For my purposes here, I'll lump these two together (as in, e.g., White 2014), but see Barnett (2019) for discussion of important differences between them.

⁷ This idea traces back to Armstrong (1968) and is defended more recently by Byrne (2009).

The knowledge first picture instead takes cases of successful perceiving as it's starting point for understanding the nature of perception. On this picture, successfully perceiving that P is one particular way of knowing that P—i.e., successful perceiving *is* a kind of knowing, a factive mental state and propositional attitude which relates a subject to her environment.⁸ In the good case, to see that there's a mug on the table, for example, is to possess a factive mental state with the content that there's a mug on the table, and it's therefore also to know that there's a mug on the table. It's important that the knowledge first view I'm describing here doesn't merely say that successfully perceiving that P *entails* knowing that P. While this entailment does hold, it holds because states of successful perceiving literally *are* states of knowing, not merely because every time one successfully perceives that P one comes to know that P.⁹ The mental state of knowledge is therefore essential to developing a full understanding of the nature of perception.

Furthermore, the knowledge first picture conceptualizes experience in the bad case as a state that aims, but fails, at putting one into a knowledge-constituting relation to the world. The bad case can be understood as a kind of 'failed perceiving,' i.e., a failure to establish a factive connection between a subject and the world. Furthermore, there's a genuinely mental difference between a subject who successfully perceives that P and her twin who hallucinates that P: in the good case, one's perceptual state is a particular way of occupying the mental state of knowledge; in the bad case, one has an experience that fails to constitute this mental state.¹⁰

⁸ For views of this sort, see Williamson (2000) and Nagel (2013; 2021). I'm putting things a bit differently from these theorists here by referring to *successful* perceiving as a way of knowing, since others often write about *genuine* perceiving as a way of knowing—i.e., they hold that genuine perception is a state of knowledge and that, in the bad case, one doesn't genuinely perceive. However, I take it that this is largely a terminological point, and that my talk of successful perceiving is more neutral between the knowledge first and internalist pictures (since both sides can agree that cases where perception provides knowledge are the cases where everything goes well perceptually). The key disagreement between them is whether, in the good case, perceiving literally is a way of knowing (where knowing is itself a mental state) or is merely a state which provides knowledge (where knowledge isn't a mental state).

⁹ See Cassam (2007) for more detailed discussion of this point. (Cassam himself adopts a 'mere entailment' view of the relationship between successful perceiving and knowing, but he helpfully positions his view in relation to the stronger, Williamson-inspired one I'm adopting.)

¹⁰ It's natural to ask whether the knowledge first picture is a kind of 'disjunctivism' about perception, though it's difficult to give a straightforward answer due to the diversity of views under the disjunctivist label (see Soteriou 2016 for a detailed overview). If disjunctivism is merely taken as

To sum up this subsection, there are two key ways in which the knowledge first picture departs from the internalist one. The first concerns the good case. For the internalist, all perceptual experiences, whether successful or not, are either experiential seemings or belief-like states. While successful perceiving does yield knowledge, this isn't a mental fact about one's experience—i.e., knowledge isn't relevant to understanding the intrinsic nature of perceiving *qua* mental state. In contrast, the knowledge first picture says successful perceiving is always a way of occupying the mental state of knowing, which means knowledge is part of a full understanding of the nature of perception. The second key departure from the internalist picture concerns the bad case. For the internalist, a subject in the bad case is mentally identical to a subject in the good case when everything experientially appears the same for each. For the knowledge first picture, the fact that an experience fails to put a subject into a knowledge-constituting relation to the world is relevant for understanding the very nature of that experience—part of what it is to hallucinate, for example, is to undergo an experience that fails to constitute knowledge.

2.3. *The (sub-)personal in epistemology and PP*

In this subsection, I'll address a preliminary worry that helps clarify my project in the rest of the paper. This worry is based on an argument that PP is inconsistent with *both* the knowledge first and internalist pictures of perceptual epistemology. This would make my project in this paper—to show that accepting PP supports the knowledge first picture over the internalist one—a non-starter.

It might initially seem that the probabilistic representations PP posits are inconsistent with the kinds of mental states posited by both of these epistemological pictures. Both the knowledge first and internalist pictures typically accept that belief is necessary for knowledge: knowledge entails belief on the knowledge first picture, while for the internalist knowledge just is belief plus appropriate further, non-mental conditions. Furthermore, though, the kind of belief necessary for knowledge is typically taken to be not merely a high degree of confidence but 'full' or 'outright' belief, which is binary: one

the idea that there's a mental difference between subjects in the good and bad cases, then the knowledge first view is disjunctivist. However, some disjunctivists hold the stronger view that there's *nothing* mentally in common between the good and bad case. The knowledge first view isn't committed to this, since it can accept that a subject who knowledgeably perceives that P and one who hallucinates that P both occupy the mental state of *seeming to perceive that P* (see Williamson 2000, ch. 1). Furthermore, many disjunctivists are committed to the 'naïve realist' idea that perception is *metaphysically direct* in the good case. The knowledge first picture is neutral about this: what matters for the knowledge first picture is whether the connection successful perception establishes between a subject and the world is a knowledge-constituting one, not whether this connection is direct or indirect.

either believes or doesn't believe that P (see, e.g., Williamson 2005; Fantl and McGrath 2009, ch. 5; Nagel 2010). Something similar goes for perceptual states, at least on the knowledge first picture: because successful perceiving is a way of occupying the propositional attitude of knowing, it involves taking the world to be determinately some way. However, PP views the brain as representing the world in a probabilistic way, employing hypotheses assigned likelihoods between 0 and 1. It might therefore seem that, if we accept PP, we have to jettison notions like outright belief and knowledge from our picture of the mind, making PP inconsistent with both the internalist and knowledge first pictures.

It would be difficult to conclusively settle here the relationship between the probabilistic processing PP describes and states like outright belief, knowledge, and perceptual experience, but I'll respond to this worry by explaining what I take to be the most plausible interpretation of their relationship. (This explanation is meant to be neutral between the knowledge first and internalist pictures, given that I aim to diffuse the worry that PP is inconsistent with both.) On my favoured interpretation, the probabilistic processing PP describes is sub-personal: it's a description of what *my brain* does, not of *my* personal-level epistemic and experiential states. One clue to this is the fact that the processing PP describes is consciously inaccessible—the 'inferences' performed by the brain in perception, for example, are automatic and sub-personal, not inferences consciously performed by the subject. There's also an obvious gap between the way PP says our brains probabilistically represent the world and the way the world seems to us. Lu et. al (2016) and Clark (2018) emphasize this gap regarding perception: in perception, what we experience is not a 'Bayesian blur' that shows the brain's leading set of hypotheses along with fainter images of competing, less likely hypotheses. Instead, we experience the world in a determinate, coherent way, in accordance with the single set of hypotheses deemed most likely (typically, at least—see Clark 2018 for discussion of perceptual phenomena in which this determinacy may break down).

Knowledge, outright belief, and perceptual experience are personal-level states, so we can retain such states in our picture of the mind by distinguishing how things are represented at the personal level from how things are represented sub-personally. In perception, this would mean having a determinate, personal-level representation with the same contents as the brain's most likely set of perceptual hypotheses, the ones that determine what we consciously experience. This can be the case even if, 'behind the scenes,' the brain is treating these as merely the most likely hypotheses among many. On this reading of PP, it describes how determinate, personal-level states like perceptual

experience are determined by sub-personal processing.¹¹ It's thus consistent with PP that we adopt an epistemological framework that includes such personal-level states, as both the internalist and knowledge first pictures do.

This helps clarify the goal of the next two sections' arguments. My goal isn't to argue that knowledge comes 'first' before the kinds of sub-personal states PP posits—sensory stimulations, probabilistic predictions, and so on—since this would be to extend a theory about personal-level states to sub-personal processes. Instead, my goal is to argue that PP's story about the relation between the sub-personal and personal levels supports the knowledge first picture of the personal-level.

3. The good case

§2.2's first key difference between the internalist and knowledge first pictures of perception concerned the good case. For the internalist, successful perceiving is either an experiential seeming or a belief-like propositional attitude; for the knowledge first theorist, successful perceiving is always a way of occupying the mental state of knowing. In this section, I consider the nature of the good case under PP in relation to these two pictures of perception.

§3.1 first argues that PP makes perceiving a propositional attitude about the world. This makes it inconsistent with the 'seemings' version of internalism, though still consistent with either the knowledge first picture or the version of internalism which says perceiving is a belief-like propositional attitude. §3.2 then argues that, under PP, one's perceptual state in the good case necessarily exhibits a characteristic necessary condition on knowledge, its 'modal robustness.' Importantly, this is so because successfully perceiving that P results from literal processes of ruling out modally nearby not-P possibilities. §3.3 then argues that, because these processes are internal to the psychology of perception, they make modal robustness a *mental* property of successfully perceiving that P. Thus, PP makes it a mental fact about one's perceptual state that it has a core, characteristic property of knowledge, which I argue supports the knowledge first picture over the internalist one.

¹¹ Whether or not all proponents of PP would accept this reading, it's at least not idiosyncratic. It's suggested by, for example, various remarks in Clark (2016), who distinguishes between 'non-conscious ('sub-personal') prediction and the shape and flow of personal-level daily experience' (78). I don't mean here to deny that we also have conscious, personal-level credences in addition to outright beliefs (cf. Weisberg 2020), but, on my reading of PP, these would also be personal-level states that result from probabilistic sub-personal processing. An alternative to my favoured picture is that all personal-level beliefs and perceptual experiences inherit graded content from the probabilistic processing that produces them. This would require accepting that perceptual experiences have graded content, which, though accepted by some (e.g., Morrison 2016), I take to be the non-standard view.

3.1. *Perceiving as propositional attitude about the world*

First, consider why PP makes perceiving a propositional attitude about the world. Proponents of PP argue that the framework blurs the traditional line between perceptual representations and ‘cognitive’ states like belief and knowledge (cf. Clark 2013, 190; Hohwy 2013, 72). That’s because, according to PP, perceptual processing involves inferences to hypotheses about the states of the world, where the ‘winning’ hypotheses are those that minimize prediction error. Because perceiving is a state which itself involves settling on an interpretation of how the world is, a process of coming to take the world to be some way is built right into the process of coming to perceive. Perceiving that P is therefore a mental state which involves taking it to be the case that P. To see a mug on the table, for example, is to settle on a hypothesis with the content that there’s a mug on the table, or to take the world to be such that there’s a mug on the table.¹²

Importantly, PP makes perceiving an attitude *about the world*. As per my explanation in §2.1, PP has it that we’re not conscious of our sensory stimulation or of a stream sense data on which we subsequently base beliefs. Rather, sensory stimulation is used, at the sub-personal level, to select a most likely hypothesis about what’s causing that stimulation. We then experience the contents of that hypothesis—i.e., we’re not conscious of sensory stimulation itself, but (in the good case) of the worldly causes of our sensory stimulation. It’s therefore not the case that successful perceiving makes us most immediately aware of how things appear or seem to us, but that it makes us aware of the world. This isn’t to deny that perception has a certain phenomenology or that there’s a way it feels to be aware of the objects causing one’s sensory stimulation. Rather, my point here is that one’s perceptual state is already an attitude *about* those objects since it results from an inference to hypotheses about them. As Clark (2016, ch. 6) explains, PP starkly contrasts with views on which perceptual experience consists in a pre-conceptual array of low-level properties on which we subsequently base distinct states of belief. In that kind of picture, when one sees a mug on the table, one’s experience consists in an array

¹² My descriptions here might seem to over-intellectualize the notion of prediction involved in PP. It’s often argued that we should think of the generative model as a kind of *structural* representation, more akin to a map of the world than a set of sentences describing the world (Gładziejewski 2016; Wiese 2017; Williams 2018). By saying perceiving that P is a propositional attitude, am I committing to a conception of the brain’s predictions as having a more complex, linguistic structure? Not necessarily: we often describe and model non-linguistic representations propositionally (when describing the contents of pictures, episodic memory, sensory imagination, etc.). Assuming it makes sense to do so, we can do the same for perception in PP. For general discussion of the relationship between non-linguistic mental representations and propositional attitudes, see Langland-Hassan (2011).

of colours, textures, shapes, and so on, and one doesn't deploy the concept *mug* or *table* until one forms a belief based on one's experience. But in PP, the world is already conceptualized in perception itself since perceiving involves the hypothesis that there's a mug on the table.

The knowledge first picture says that successful perceiving is a way of knowing about the world, where knowledge is a propositional attitude. So, PP coheres with the knowledge first picture in that both make perceiving a propositional attitude about the world. This means PP also excludes the kind of internalist view on which perceptual experience is merely a source of appearances or seemings on which we subsequently base beliefs. On that picture, we come to take the world to be some way by forming beliefs which are distinct states from states of experience. On the PP picture, there's no prior state of experience which then causes a distinct state of belief, so this version of internalism assumes a picture of the psychology of perceptual belief-formation which PP rules out.

However, everything I've just argued is still consistent with an internalist picture on which successful perceiving is a kind of belief or belief-like state rather than constituting knowledge. To shore up the knowledge first picture over this internalist one, it also matters whether PP makes perceiving that P a way of knowing, where this is a mental fact about the nature of perceiving. The next two subsections build towards this conclusion.

3.2. *Successful perceiving as modally robust*

This subsection first identifies a core, distinctive feature of knowledge: that it's necessarily 'modally robust.' I then argue that this same necessary condition on knowledge applies to successful perceiving under PP, in virtue of how perceptual processing involves various ways of ruling out alternative possibilities.

It's uncontroversial that knowledge requires truth, but epistemologists also typically take knowledge to require more than merely getting things right: to know that P is to be right about P in a way that reliably tracks the facts. This means that the way one represents the world is disposed to vary with the way the world itself varies, such that, had the world been slightly different, one would still have accurately represented it. This property of knowledge is often cashed out in terms of knowledge's modal profile. There are various precise ways of formulating modal robustness conditions on knowledge, but the kind I'll focus on is the *relevant alternatives* formulation. According to relevant alternatives views, to know that P is to be in a state which rules out that one is in the set of relevant not-P possible worlds, or the set of relevant possible worlds in which P is false. Exactly what counts as a 'relevant' possibility is controversial, but such accounts typically take far-fetched Cartesian

skeptical scenarios to be irrelevant and take nearby possibilities that closely resemble the actual world to be relevant.¹³ Modal robustness conditions of this sort are a widespread, core feature of many accounts of knowledge, both internalist and knowledge first accounts.¹⁴

Now, before turning specifically to modal robustness as a *necessary* condition on successful perceiving, consider how in general PP builds processes of ruling out nearby possibilities right into perceptual processing. For one thing, PP says perceiving involves selecting one hypothesis as most probable relative to competitors, using incoming sensory signals to settle on the most likely hypothesis. By settling on a single hypothesis, the brain thereby uses cues from the world around oneself as a basis for ruling out alternative, competing hypotheses about the states of that world. Specifically, this involves ruling out competing hypotheses that are modally *nearby* the winning hypothesis, in virtue of the brain's reliance on abstract prior expectations to constrain the space of entertained alternative hypotheses. Recall, from §2.1, that such priors are very high-level expectations that constrain the brain's hypothesis space, expectations about matters like object permanence and the physics of object interactions. The result is that the set of competitor hypotheses, against which one hypothesis is deemed most probable, aren't far-fetched possible worlds in which, for example, objects randomly pop in and out of existence or Cartesian skeptical scenarios obtain. Instead, the brain rules out competitor hypotheses which are close to the actual world, candidates for how the world could actually have been.¹⁵

Active environmental sampling, the kind of 'epistemic action' I described in §2.1, is another sense in which PP incorporates the ruling out of nearby alternatives into perceptual processing. Such actions involve interacting with the world to sample sensory data that helps to (dis)confirm our current leading perceptual hypothesis. The brain generates expectations, under its current leading hypothesis,

¹³ For classic relevant alternatives accounts, see Stine (1976) and Goldman (1976). Theorists disagree about whether, and exactly how, farther-out possibilities become relevant as one moves between different conversational contexts—for example, whether merely entertaining a possibility makes it relevant (cf. Lewis 1996) or whether the practical importance of knowing P changes which possibilities are relevant (cf. Cohen 1999).

¹⁴ For internalists, this is because such modal robustness is a condition beliefs must meet to amount to knowledge, in addition to other conditions like truth. For knowledge first proponents, it's because knowledge is a mental state which essentially or necessarily links one to truths, rather than one which merely happens to involve truth.

¹⁵ This is not to say that each particular brain's ordering of the likelihoods of these alternative possibilities reflects their exact ordering in modal space, besides the fact that they're modally nearby. What matters for the modal robustness of perception is whether alternative possibilities are in fact ruled out, not whether the probabilistic process by which they're ruled out precisely models their modal distribution.

about how the sensory signal will evolve as one acts on the world; as these expectations are confirmed, the probability of competing hypotheses is reduced. That's because, if the sensory signal *had* evolved in line with a competing hypothesis, the current leading hypothesis would have to be revised to an alternative one that better fits the sampled sensory data.¹⁶ So, active sampling is a matter of confirming one's leading hypothesis against competing ones; it's used to confirm that the brain's leading hypothesis reflects the way the world actually is, rather than the world being some different, modally nearby way.

Now, if what the brain infers when we perceive the world is really a probability distribution over competing hypotheses, it might seem wrong to describe this as a process of *ruling out* alternatives—isn't it really the case that alternatives are not altogether ruled out but treated as low-probability-yet-still-possible? This may be right when it comes to how the brain's sub-personal processing represents things. However, we should keep in mind that what we experience at the personal level are the determinate contents of a single hypothesis, not a probability distribution (as per §2.3 above). The brain inferring a distribution, with a single, winning hypothesis at its peak, is therefore a means of ruling out competing hypotheses when it comes to determining one's personal-level perceptual state. So, perceptual processing effectively involves rejecting not-P possibilities as candidates for determining the contents of one's personal-level perceptual state.

So, on the PP account of perception, perceptual processing involves ruling out nearby alternative possibilities. I'll now argue that it's *necessary* for successfully perceiving that P that perceptual processing rule out not-P possibilities. That's because cases in which such ruling out fails to occur are instances of bad cases.¹⁷

Before considering specifically modally *nearby* possibilities, first consider how an experience which comes about in absence of *any* ruling-out process whatsoever results in a bad case. We can imagine a hypothetical case in which the brain's top-down predictions wholly determine what one

¹⁶ Or, one would have to change the world to bring it in line with one's current leading hypothesis. For simplicity, I've left this kind of 'active inference' out of my account of PP in this paper, though it's an important missing component (see Clark 2016, ch. 4). Regardless, this process would still help to ensure the modal robustness of one's winning hypothesis, because it rules out an alternative to this hypothesis—this alternative being the way the world currently is, which gets 'ruled out' by changing the world.

¹⁷ This isn't to say *all* possible bad cases result from a failure to rule out modally nearby possibilities. I claim that such failure is sufficient for being in a bad case, but there may also be other routes into bad cases. This section focuses on bad cases that illustrate that it's necessary for the good to rule out modally nearby hypotheses, but §4 discusses other bad cases.

experiences, due to a malfunction at some stage of the bottom-up components of perceptual processing (taking into account sensory data, generating prediction error, and/or revising hypotheses on the basis of prediction error). If this were to occur, the resulting experiential state would seem more like a hallucination than successful perception, because it would be an experience generated entirely internally rather than one guided by the way things are in the world. (This is somewhat of an oversimplified, hypothetical case, and PP accounts of how actual hallucinations work are more complex—I come back to this below in §4). In this case, one’s perceptual processing goes awry precisely because the brain’s hypothesis-selection process failed to be guided by bottom-up sensory signals—i.e., because the brain failed to use bottom-up sensory signals to rule one hypothesis most likely relative to alternatives, instead just settling on a winning hypothesis without engaging in the typical ruling-out process.

Furthermore, it’s plausibly also necessary for maintaining our conscious perceptual grip on the world that we continuously engage in ruling out alternatives via active environmental sampling. This is because our generative models include the abstract expectation that the world around us is dynamic and liable to change over time. As such, there’s always some degree of uncertainty about whether some part of the world from which we’ve recently received sensory stimulation has since changed, uncertainty which accumulates over time (Hohwy, Paton, and Palmer 2016). This explains why we naturally go about actively observing and checking our environment as we interact with it, through means like eye saccades. Parr et al. (2019) argue that, furthermore, such sampling actually sustains the veridicality of our perceptual representations, by blocking the accumulation of uncertainty that would cause the brain’s current leading perceptual hypotheses to be deemed less and less likely over time. They ground this claim in an appeal to the phenomenon of ‘Troxler fading,’ in which forcing oneself to fixate on the center of an image causes more peripheral parts of one’s visual experience to fade into a uniform grey field. This phenomenon suggests that artificially blocking our usual saccadic environmental sampling causes the fidelity of our visual representation of the world to deteriorate, because uncertainty about the parts of the environment on which we’re not fixating accumulates. If this argument is right, it’s another sense in which successfully perceiving that P depends on processes that rule out alternative possibilities: maintaining a veridical perceptual representation requires that we actively continue to rule out alternatives.

So, on the PP framework, the good case requires that perceptual processing rule out alternative, not-P possibilities. Moreover, it plausibly also requires that, specifically, *nearby* not-P

possibilities are ruled out. That's because, when the brain rules out some possibilities but fails to rule out nearby ones, it results in a bad case.

To see this, we can imagine a hypothetical case in which, despite settling on a true winning hypothesis P, the brain rules out only far-fetched not-P possibilities rather than modally nearby ones, because it has counted only far-fetched possibilities as likely competitors to this winning hypothesis. Suppose that one is about to walk into the kitchen where there's a mug on the table, and that the brain generates a probability distribution which has at its peak the true hypothesis that there's a mug on the table. However, rather than distributing lower probability values over a set of hypotheses that includes modally nearby possibilities (e.g., that there's a table but no mug; that there's a plate on the table instead of a mug; etc.), the brain's only alternative hypotheses are quite far-fetched: that there's a unicorn on the table, that there's a blackhole in the room instead of a table, and so on. Nevertheless, because the initial leading hypothesis successfully explains the brain's sensory stimulation and minimizes prediction error, the experience that results from deeming this hypothesis successful is veridical.

Despite the resulting veridical experience, it seems intuitively that something has gone wrong in this case: namely, the brain would not have succeeded in veridically tracking the states of the world if one had been in a nearby possible world where there was no mug on the table, because the brain entertained a possibility space that didn't include any such nearby worlds. In other words, although the brain happened to get things right, it does so in a way that's not disposed to track facts about the world in a modally robust way, since things would have malfunctioned if the world had been slightly different. In virtue of this, it seems like the brain's perceptual processing has been unsuccessful in this case. The idea that modal robustness is a necessary condition on successful perceiving has a tidy explanation for this intuition: one's experiential state is sub-optimal because it failed to be modally robust across nearby possible worlds.

3.3. Modal robustness as mental fact: why PP supports the knowledge first picture

PP's picture of perceptual processing is thus one in which successfully perceiving that P results from processes of ruling out nearby not-P possibilities. Successful perceiving is therefore a state which possesses the kind of modal robustness that's characteristic of knowledge. I now argue that, furthermore, PP makes modal robustness a *mental* feature of successful perceiving and that this supports the knowledge first picture of the good case over internalism.

Under PP, it's not merely that to successfully perceive the world is to be in a state which also entails that one is not in a nearby not-P possibility. Instead, PP makes this ruling-out process internal to the psychology of perception: coming to successfully perceive the world literally *is* a process in which the brain represents and rules out nearby alternatives. It's therefore an intrinsic fact about perceptual processing that it aims at putting one in touch with one's environment *by* ruling out alternatives, where cases of successful perceiving are those in which this is accomplished. When it does this successfully, the resulting state is one which hits on the truth about one's environment in a modally robust way. So, once we adopt the PP picture, it should become part of our understanding of successful perceiving as a mental state that it involves this kind of modal robustness.

We can sharpen this picture by contrasting it with one on which successful perceiving merely *entails* modal robustness or is always accompanied by it. In fact, this kind of picture emerges from a more traditional, bottom-up account of perception, on which successful perceiving involves detecting features of one's environment as presented in sensory stimulation. On this 'detection' model, one's perceptual states are robust across nearby possible worlds when one's perceptual systems are sensitive to how things are in one's environment. This would be so as long as how things seem in perceptual experience is disposed to vary with how things are in the world: if it *had* been the case that some not-P possibility Q were the case, then one would not have experienced P because one's perceptual system would be sensitive to the fact that Q rather than P. On this picture, though, the modal robustness of perception is grounded in the truth of counterfactuals about what one would perceive were one in nearby possible worlds. Such counterfactual truths are 'external' to a subject's mind, so we don't necessarily need to bring modal robustness into the picture in order to understand the intrinsic nature of perceiving as a mental state. On this bottom-up account, it would be incorrect to say that successful perceiving *is*, by its very nature, a mental state of getting things right by ruling out alternatives; instead, successful perceiving would be a state of detecting facts about one's environment, which also entails that nearby alternatives are ruled out. So, on this bottom-up picture, it seems more plausible than on the PP picture to say that successful perceiving merely entails that not-P possibilities are ruled out, rather than that such ruling out is intrinsic to successful perceiving.

One might object that I've just conflated psychological facts about the processing that underlies perception with metaphysical facts about the mental state of perceiving. Even if it's true that ruling out alternative possibilities is part of the sub-personal psychological processing that *produces* a successful, personal-level perceptual state, one might argue that these psychological facts have no bearing on the intrinsic nature of that state itself. However, as Williamson (2000, 41) notes, it seems

like a general feature of mental states that facts about their psychological, causal origins bear on their intrinsic natures. This applies most clearly to how we individuate types of mental states—for example, we can differentiate between remembering that P and merely imagining that P based at least in part on the causal history of one's mental representation. Similarly, whether or not some judgment counts as an inference depends on whether the process which produced it was an inferential one. In these cases, the psychological processes that produce a personal-level mental state clearly have implications for metaphysical questions about the intrinsic nature of that state. Analogously, I've argued that, when PP makes a process of ruling out alternatives internal to the psychological processing that underlies successful perceiving, this has metaphysical implications for the nature of the resulting mental state.

I've argued in this section so far that, under PP, it's a mental fact that successfully perceiving the world is a propositional attitude which has the kind of modal robustness that's characteristic of knowledge. This means that it's a mental fact about successful perceiving that it *at least closely resembles* knowledge, in virtue of exhibiting one of knowledge's core properties. This gets us close to the knowledge first picture's claim that it's a mental fact that successful perceiving *is* a way of knowing, so it gives us positive reason to accept this claim. In other words, the idea that it's a mental fact about successful perceiving that it has one of the core properties of knowledge is good reason to think successful perceiving is a way of knowing, where this is a mental fact about successful perceiving.

Of course, this isn't an airtight entailment from PP to the knowledge first account of successful perceiving. Knowledge has other properties besides modal robustness—as Williamson (2000) argues, there seems to be an open-ended list of necessary conditions one must meet to have knowledge (e.g., possessing no defeaters, not being 'Gettiered,' etc.). I haven't canvassed all of these to show that PP implies successful perceiving has these properties. However, if we want at least a succinct *approximate* definition of knowledge, it's difficult to find a better description than 'hitting on the truth that P in a modally robust way.' Since it's fundamental to the nature of knowledge that it's a factive state, knowing that P involves *necessarily* being right about P (rather than merely happening to be right about P in a way that could've easily gone wrong). So, it's particularly central to knowledge that it involves modal robustness. Showing that this central property of knowledge is a mental feature of successful perceiving is thus good evidence that successful perceiving constitutes the mental state of knowledge.

Now, the internalist agrees that successful perceiving is a source of knowledge and therefore predicts that successful perceiving entails the modal robustness of knowledge. However, the internalist doesn't hold that successful perceiving *is* a way of knowing, where this is a characterization of the intrinsic nature of successful perceiving as a mental state. Instead, internalism says that knowledge is

constituted by a combination of mental and non-mental factors: the mental state of belief plus non-mental factors like modal robustness. So, recognizing the way modal robustness is built right into successful perceiving as a mental state provides more support for the knowledge first picture than the internalist one. Rather than successful perceiving yielding knowledge because it merely entails a non-mental property of modal robustness, knowledge-like modal robustness is a mental fact about successful perceiving.

I'll close this section with a more general line of argument. Under PP, successful perceiving results directly from processes that, from the perspective of traditional epistemology, are closely associated with knowledge. Our theory of perceptual epistemology would be missing out on epistemologically interesting features of successful perceiving if we didn't take this into account. Adopting a knowledge first framework does this for the reasons I've covered in this subsection. Furthermore, the features of perceptual psychology under PP help to flesh out the knowledge first picture and *explain why* it's true. PP shows how, in virtue of features internal to perceptual processing, successful perceiving gains the modal robustness characteristic of knowledge. It thereby provides a partial explanation of why successful perceiving is a way of knowing, where this is grounded in psychological features of perceptual processing itself.

So, adopting both PP and the knowledge first picture together results in a unified theoretical account of perceptual epistemology and psychology. Adopting an internalist picture yields a less unified account, since the internalist doesn't take into account the fact that the knowledge-like properties of successful perceiving are genuinely mental properties.

I therefore conclude in this section that the account PP gives of the good case lends support to the knowledge first picture over internalism. Of course, the knowledge first picture is broader than just how it characterizes the good case. I turn to considering the bad case in the next section.

4. The bad case

The following line of thought is natural in response to the previous section's arguments. Perhaps it's true that, in the good case, perceiving is a propositional attitude that looks a lot like knowledge, where this is a mental fact about perceiving. But the knowledge first picture also says something about how to conceptualize the bad case. On this picture, any experience which fails to constitute knowledge is unsuccessful; furthermore, there's a mental difference between subjects in the good and bad cases, since subjects in the bad case lack the mental state of knowledge. However, In PP, all experiences result from the same material: the top-down expectations the brain has deemed

the most likely. Doesn't this mean that, contrary to the knowledge first picture, there's no real mental difference between the nature of experiences that do and don't constitute knowledge?

This section argues against this line of thought. §4.1 first considers several kinds of bad cases I touched on in the previous section, showing how PP encourages a conceptualization of these cases that supports the knowledge first picture. §4.2 considers PP treatments of perceptual illusions and argues that we should construe these in a way that supports the knowledge first picture, while §4.3 does the same for hallucinations. §4.4 then considers an additional kind of bad case which is difficult to make sense of under the knowledge first picture: cases where one's experience results from a deviant causal chain. I show how PP helps flesh out a knowledge first treatment of such a case.¹⁸

4.1. Failures to rule out modally nearby hypotheses

In the previous section, I touched on three kinds of unsuccessful perceiving: cases in which one's experience is determined by top-down predictions not revised at all against bottom-up sensory signals; cases in which, as in the phenomenon of Troxler fading, one's veridical experience deteriorates due to the accumulation of uncertainty that would normally be blocked by epistemic actions like eye saccades; and cases in which one's brain settles on a true perceptual hypothesis as most likely but fails to rule out nearby alternative hypotheses, instead ruling out only far-fetched ones.

Given the arguments of the previous section, we can give a unifying explanation of what goes wrong in each of these unsuccessful cases: all are cases in which perceptual processing goes awry because nearby alternatives aren't ruled out. It's not merely that such cases fail to *entail* that alternative hypotheses are ruled out. Rather, such cases involve a perceptual state produced via a process that literally fails to rule out nearby alternatives. The resulting mental state fails to be modally robust in the right way, where this is a property of the mental state itself rather than something merely entailed by the mental state.

¹⁸ There's one prominent kind of bad case I'm not touching on in this section: Cartesian skeptical scenarios like the brain-in-a-vat. Such scenarios might seem especially tricky for the knowledge first picture, since they involve a brain that, though it lacks knowledge, seems psychologically identical to its good case twin. However, my view is that PP is particularly conducive to the arguments of Chalmers (2005), according to which brain-in-the-vat-type scenarios aren't actually bad cases after all; see Hohwy (2017) and Clark (2017) for discussion of this point. Unfortunately, discussion of these large issues would take me too far afield here. But adopting a Chalmers-style assessment of brain-in-a-vat cases allows us to hold both that the brain-in-a-vat and her good case twin are mentally identical *and* that both have perceptual knowledge, meaning such cases pose no counterexample to the knowledge first picture.

Similar to what I argued about the good case in the previous section, this assessment of these cases of unsuccessful perceiving supports the knowledge first picture. In these cases, it's a mental fact that one's experiential state is unsuccessful because it lacks the modal robustness that's a core property of knowledge. This doesn't get us all the way to the knowledge first picture of such cases, which would simply say it's a mental fact that such experiences are unsuccessful insofar as they aren't instances of knowledge. But I've at least argued that it's a mental fact that such experiences are unsuccessful because they lack a *core property* of knowledge, which gives us positive reason to think it's a mental fact that they're unsuccessful because they fail to *be* knowledge.

Internalism would agree that these bad cases are unsuccessful because they don't yield knowledge. However, it denies that this is in virtue of a *mental* difference between the good case and bad case. So, adopting a PP account of these bad cases gives us more reason to accept the knowledge first picture than internalism.

4.2. *Perceptual illusions*

Given given PP accounts of perceptual illusions, it might seem that illusions cause problems for my arguments in the paper so far. Consider, for example, the 'hollow mask' illusion, in which a concave mask or face is perceived as convex. A plausible PP explanation of this illusion is that the generative model includes a deeply entrenched expectation that faces are convex, which can't be overridden even if one knows the face one is looking at is concave (Frith 2007, 127-30; Wilkinson 2014, 147). This prior expectation can't be revised because doing so would be detrimental to overall prediction error in the long run: because faces in naturalistic settings almost always *are* convex, the best strategy to minimize overall prediction error is to not allow this prior to be revised, despite special cases where it fails (Clark 2016, sec. 6.11). It therefore might seem wrong to call this a kind of 'failed perception,' which is how the knowledge first picture treats all bad cases. That's because one might think nothing has gone *wrong* here if one is employing a generative model that's optimally configured for minimizing overall prediction error.

However, Hohwy (2013, ch. 6) develops an account of illusions which allows them to be absorbed into my account in this paper. Hohwy argues that perceptual illusions occur when one is artificially restrained from carrying out epistemic actions, i.e., prevented from actively interacting with one's environment to test predictions about how one's actions would effect it. Recall that one's generative model encodes expectations about how the world would evolve if one were to interact with it, and therefore that acting on the world is one way to test whether one's dominant perceptual

hypothesis is accurate. Hohwy points out that, in many cases, one can disrupt illusions by actively manipulating them. This goes for the hollow mask case, where one can disrupt the illusion by touching the mask rather than passively looking at it. So, it seems the illusion occurs when one is unable to actively test one's leading hypothesis that the mask is convex.¹⁹

This treatment of illusions suggests that they emerge when one is unable to rule out false hypotheses via epistemic actions. This makes illusions another kind of case in which one's perceptual experience fails to constitute knowledge because it fails to be modally robust. So, we can treat such cases in the same way as those I discussed in the previous subsection: illusions are a failure to achieve a knowledge-constituting perceptual state in the sense that one hasn't ruled out false possibilities in the way that's necessary for knowledge.

4.3. Hallucinations

Hallucinations don't fit nicely with the bad cases I've discussed so far, since they don't necessarily result from failures to rule out modally *nearby* possibilities—a hallucination with some wild, far-fetched content may instead result from failure to rule out *far away* hypotheses. Nevertheless, this subsection argues that adopting a PP account of hallucinations fits well with the knowledge first picture.

It's controversial exactly how to explain hallucinations within PP, but I'll focus on one prominent kind of account, according to which hallucinations occur when persistent prediction errors are *wrongly generated* (Fletcher and Frith 2009; Clark 2016, sec. 2.12).²⁰ Suppose your brain correctly predicts that there's a mug on the table and receives sensory stimulation that confirms this. However, due to malfunctioning prediction error mechanisms, prediction error signals continue to be generated despite your correct prediction. This would force your brain to continue revising its model of the world to accommodate these error signals, even though you originally got things right. This might even require the brain to find a model of the world that departs from reality. When it then generates

¹⁹ Hohwy himself doesn't discuss the hollow mask example but applies his account to other illusions: the Müller-Lyer illusion (where manipulating the image in the right way allows one to veridically see the lines as equal) and multisensory illusions like the rubber hand (where the illusion persists only under conditions in which bodily movement is suppressed).

²⁰ There are also other PP accounts of hallucination (e.g., Wilkinson 2014; Benrimoh et al. 2018); what I argue in this subsection should be adaptable to them.

new perceptual predictions based on this reality-incongruent model, the result would be a hallucinatory experience.²¹

Now, PP literature sometimes characterizes (successful) perception as ‘controlled hallucination,’ since it occurs in a top-down, internally generated way that’s controlled by the world via bottom-up sensory stimulation. However, following Clark (2016, sec. 6.10), I think it makes more sense to call hallucination a kind of *uncontrolled perception*. The label ‘controlled hallucination’ seems to imply that there’s a prior category of hallucination in terms of which we should understand the nature of successful perception, with the latter a sub-kind of the former. However, the above PP account instead makes hallucinations disruptions to a system whose normal function is to be controlled by the world. So, we should take the category of successful perceiving that’s controlled by the world to be primary and understand hallucinations as cases in which perceptual machinery fails to be properly controlled.²²

§3 argued that, in the good case, successful perceiving is a way of knowing. So, to say that hallucination is a form of failed perception is akin to saying that hallucinations are a failure to produce knowledge. This fits well with the knowledge first idea that we should conceptualize bad cases as cases that, by their very nature, fall short of achieving knowledge. And it fits poorly with the internalist picture, which treats successfully perceiving that P and hallucinating that P as mentally on par.

4.4. *Deviant causal chains*

I’ll now consider another kind of bad case: one in which the sensory stimulations one receives from the world are produced via a deviant causal chain.²³ My argument regarding such cases will have a slightly different structure than my main arguments in the paper so far, which have focused on how

²¹ This is a very simplified account of hallucination, but it’s sufficient for my purposes. One obvious thing I’ve left out is *why* persistent, false error signals might be generated in the first place; see again Fletcher and Frith (2009). I’ve also overlooked the importance for PP accounts of hallucination of a mechanism known as ‘precision weighting,’ by which the brain adjusts the relative weightings of top-down predictions and bottom-up error signals.

²² The idea that hallucination is a kind of unsuccessful perception is especially stark in PP versus a bottom-up account of perception. On the latter, since perception involves receiving external sensory stimulation and extracting information from it, it’s not totally clear what the relationship between perception and hallucination is—hallucinations don’t seem very perception-like, given that they’re generated internally (cf. Wilkinson 2014, 149). At the very least, it’s not clear whether to think of hallucination as an unsuccessful version of perception, given their different aetiologies. Under PP, it’s more straightforward to see hallucination as unsuccessful perception, since it involves the same top-down predictions as in perception failing to be properly controlled.

²³ Thank you to Jennifer Nagel for alerting me to the relevance of this sort of case.

PP provides support for the knowledge first picture. My aim in the rest of this section is to reveal another kind of coherence between the two frameworks by pointing out an interesting sense in which PP can help to enrich the knowledge first picture.

Consider the following case. Suppose there's a burning candle on the table in front of you and a mirror between you and the candle, such that, from where you're standing, you can't perceive the candle directly. However, unbeknownst to you, a chain of additional mirrors has been rigged up around the room in such a way that you perceive a reflection of the burning candle in the mirror. The reflection of the candle looks to you exactly as the candle actually does, and the candle seems to you to be in the location on the table where it in fact is. In other words, when you look at the mirror, you have a veridical experience as of the candle on the table. While this experience ultimately causally traces back to the candle, it does so via a chain of reflections in mirrors. Despite the experience's veridicality, it's typically thought that a deviant causal chain of this sort wouldn't yield knowledge that there's a candle on the table (cf. Peacocke 1979; Plantinga 1993, ch. 9).

Under the knowledge first picture, such a case would be one where the experience itself fails to constitute the mental state of knowledge. However, it's actually somewhat tricky for the knowledge first picture on its own to give an account of *why* this experience falls short of knowledge. It's tempting to think that your deviantly-caused experience is in some sense only accidentally veridical, because the candle's reflection in the mirror could have presented things non-veridically if, for example, the mirrors were arranged slightly differently. However, we could just stipulate that whoever rigged up the system of mirrors is very reliable, such that there are no nearby possible worlds in which one experiences things non-veridically. Under this stipulation, the contents of one's experience are true in a reliable, modally robust way. Still, it doesn't seem intuitively like the experience yields knowledge, in virtue of the fact that you're unknowingly standing in a direct causal relation to the candle's reflection rather than the candle itself.

However, PP helps us explain why this experience is unsuccessful relative to a non-deviant, knowledge-constituting experience. Recall, from §2.1, that PP doesn't merely say perceptual processing is a matter of settling on the most likely hypotheses about the world; PP also says, more specifically, that perceptual processing is a matter of settling on hypotheses that *model the causes of the brain's sensory stimulation*. So, any case which fails to do this successfully is one that fails to successfully carry out the task of perception as a psychological capacity. This applies to our deviant causal chain case. In that case, the brain has inferred the true hypothesis that there's a candle on the table; however, its sensory stimulation was not caused directly by the candle itself, but by the mirror. There's therefore

a sense in which, under a PP account of perception, one's brain gets things wrong here. While it's true that there's a candle on the table, one's brain has failed to successfully carry out the task of modelling the causes of one's sensory stimulation. The brain's winning hypothesis fails take into account a more proximal cause, the mirror.²⁴

PP thus reveals a sense in which this kind of experience falls short of successful perceiving. Furthermore, this construal of the case gives us a plausible account of why the experience doesn't constitute knowledge. Again, for one's brain to settle on the hypothesis that there's a candle on the table is for one's brain to pinpoint the candle as the cause of one's sensory stimulations. In our deviant causal chain case, this means that one's experience results from a process which *falsely* takes the candle to be the cause of one's experience, rather than the mirror with its reflection of the candle. This case therefore bears striking resemblance to 'false lemmas' cases in the epistemological literature—cases where one accidentally forms a true belief based on a false premise. Because they involve believing a truth based on a falsehood, such cases are typically taken not to constitute knowledge (see, e.g., Levin 2006). Now, such cases in the literature typically involve inferring a true, personal-level belief from a false, personal-level belief. In contrast, the deviant perceptual case I'm considering involves a true perceptual hypothesis that's determined, at a sub-personal level, by a process which involves a falsehood. Still, it seems intuitively plausible to extend the same kind of reasoning to this perceptual case, since it still involves taking something to be true in a way that's derived from a falsehood. In virtue of this, there seems to be something epistemically defective about the resulting perceptual state.²⁵

So, PP provides a knowledge first-friendly treatment of a kind of case that was difficult to account for within the knowledge first picture alone. In addition to all the arguments I gave above for why PP supports the knowledge first picture, my argument here shows why accepting both

²⁴ As an anonymous reviewer helpfully points out, we can make it especially clear why one's perceptual representation is unsuccessful here by considering how it fails to afford long-run prediction error minimization. The generative model encodes predictions about what would occur were one to interact with one's environment, which would include, for example, that one will feel the flame's heat if one reaches towards the candle. In this deviant case, though, reaching towards the candle will instead cause you to bump into the mirror, generating prediction error. This highlights why one's generative model is sub-optimal when it fails to represent the mirror: it's not optimized for prediction error minimization across possible interactions with one's environment. One can only resolve this by incorporating the mirror into one's model of the world.

²⁵ For arguments that states involved in sub-personal perceptual processing can stand in epistemological relations to other states, including personal ones, see Jenkin (2020).

frameworks together is theoretically fruitful. This reveals a further, interesting point of deep coherence between PP and the knowledge first picture.

6. Conclusion

I've argued that adopting the PP framework for perceptual psychology supports adopting a knowledge first understanding of the nature of perception. PP supports the knowledge first picture on two fronts: in the way it treats the good case of successful perceiving as well as the bad case of experience that fails to constitute knowledge.

Once we accept the knowledge first picture, we can do away with the misleading idea, mentioned in §1, that PP makes perceptual experience akin to a fantasy or virtual reality. It's true that PP says we perceive the world by modelling it in an internally generated, top-down way. But this modelling process is one which uses prediction error minimization to rule out 'false realities' and thereby generate a representation of the world which, as a feature of perception's internal psychological processing, has the kind of modal robustness characteristic of knowledge. There's thus nothing 'virtual' or 'fantastical' about what we perceive; rather, perception latches on to the actual world in a knowledge-constituting way, making us conscious of the external, worldly causes of our sensory stimulation.

Moreover, my arguments show that another idea with which I opened the paper—that knowledge doesn't belong in scientifically respectable theories of the mind—is a misconception. Far from it, there are actually deep affinities between a powerful lens for scientific study of the mind and the view that knowledge is key to our understanding of perception. It's therefore not a foregone conclusion that naturalistic study of the mind should focus strictly on internal states at the expense of knowledge.

Bibliography

- Armstrong, D. M. 1968. *A Materialist Theory of the Mind*. London: Routledge & Kegan Paul.
- Barnett, David James. 2019. "Perceptual Justification and the Cartesian Theatre." In *Oxford Studies in Epistemology Volume 6*, edited by Tamar Szabó Gendler and John Hawthorne, 1-34. Oxford: Oxford University Press.
- Benrimoh, David, Thomas Parr, Peter Vincent, Rick A. Adams, and Karl Friston. 2018. "Active Inference and Auditory Hallucinations." *Computational Psychiatry* 2: 183-204.
- Byrne, Alex. 2009. "Experience and Content." *The Philosophical Quarterly* 59 (236): 429-451.
- Cassam, Quassim. 2007. "Ways of Knowing." *Proceedings of the Aristotelian Society* 107 (3): 339-358.
- Chalmers, David J. 2005. "The Matrix as Metaphysics." In *Philosophers Explore the Matrix*, edited by Christopher Grau. New York: Oxford University Press.
- Churchland, Patricia Smith. 1987. "Epistemology in the Age of Neuroscience." *The Journal of Philosophy* 84 (10): 544-553.
- Clark, Andy. 2013. "Whatever Next? Predictive Brains, Situated Agents, and the Future of Cognitive Science." *Behavioral and Brain Sciences* 36 (3): 181-204.
- . 2016. *Surfing Uncertainty: Prediction, Action, and the Embodied Mind*. Oxford: Oxford University Press.
- . 2017. "How to Knit Your Own Markov Blanket." In *Philosophy and Predictive Processing*, edited by Thomas Metzinger and Wanja Wiese. Frankfurt: MIND Group.
- . 2018. "Beyond the 'Bayesian Blur': Predictive Processing and the Nature of Subjective Experience." *Journal of Consciousness Studies* 25 (3-4): 71-87.
- Clark, Andy, Karl Friston, and Sam Wilkinson. 2019. "Bayesing Qualia: Consciousness as Inference, Not Raw Datum." *Journal of Consciousness Studies* 26 (9-10): 19-33.
- Cohen, Stewart. 1999. "Contextualism, Skepticism, and the Structure of Reasons." *Philosophical Perspectives* 13: 57-89.
- Fantl, Jeremy and Matthew McGrath. 2009. *Knowledge in an Uncertain World*. Oxford: Oxford University Press.
- Fletcher, Paul C. and Chris D. Frith. 2009. "Perceiving is Believing: A Bayesian Approach to Explaining the Positive Symptoms of Schizophrenia." *Nature Reviews Neuroscience* 10 (1): 48-58.
- Fodor, Jerry. 1980. "Methodological Solipsism Considered as a Research Strategy in Cognitive Psychology." *Behavioral and Brain Sciences* 3 (1): 63-73.

- Friston, Karl. 2005. "A Theory of Cortical Responses." *Philosophical Transactions of the Royal Society B: Biological Sciences* 360 (1456): 815-836.
- Friston, Karl, Francesco Rigoli, Dimitri Ognibene, Christoph Mathys, Thomas Fitzgerald, and Giovanni Pezzulo. 2015. "Active Inference and Epistemic Value." *Cognitive Neuroscience* 6 (4): 187-214.
- Friston, Karl, Philipp Schwartenbeck, Thomas FitzGerald, Michael Moutoussis, Timothy Behrens, and Raymond J. Dolan. 2014. "The Anatomy of Choice: Dopamine and Decision-Making." *Philosophical Transactions of the Royal Society B: Biological Sciences* 369 (1655): 20130481.
- Frith, Chris. 2007. *Making Up the Mind: How the Brain Creates our Mental World*. Malden, MA: Blackwell.
- Ganeri, Jonardon. 2010. "The Study of Indian Epistemology: Questions of Method—A Reply to Matthew Dasti and Stephen H. Phillips." *Philosophy East and West* 60 (4): 541-550.
- Gładziejewski, Paweł. 2016. "Predictive Coding and Representationalism." *Synthese* 193 (2): 559-582.
- Goldman, Alvin I. 1976. "Discrimination and Perceptual Knowledge." *The Journal of Philosophy* 73 (20): 771-791.
- Hobson, J. Allan and Karl J. Friston. 2014. "Consciousness, Dreams, and Inference: The Cartesian Theatre Revisited." *Journal of Consciousness Studies* 21 (1-2): 6-32.
- Hohwy, Jakob. 2013. *The Predictive Mind*. Oxford: Oxford University Press.
- . 2016. "The Self-evidencing Brain." *Noûs* 50 (2): 259-285.
- . 2017. "How to Entrain Your Evil Demon." In *Philosophy and Predictive Processing*, edited by Thomas Metzinger and Wanja Wiese. Frankfurt: MIND Group.
- Hohwy, Jakob, Bryan Paton, and Colin Palmer. 2016. "Distrusting the Present." *Phenomenology and the Cognitive Sciences* 15 (3): 315-335.
- Huemer, Michael. 2006. "Phenomenal Conservatism and the Internalist Intuition." *American Philosophical Quarterly* 43 (2): 147-158.
- Jenkin, Zoe. 2020. "The Epistemic Role of Core Cognition." *Philosophical Review* 129 (2): 251-298.
- Jenkins, C. S. 2007. "Epistemic Norms and Natural Facts." *American Philosophical Quarterly* 44 (3): 259-272.
- Kim, Jaegwon. 1988. "What is 'Naturalized Epistemology?'" *Philosophical Perspectives* 2: 381-405.
- Kirchhoff, Michael D. 2018. "Predictive Processing, Perceiving and Imagining: Is to Perceive to Imagine, Or Something Close to it?" *Philosophical Studies* 175 (3): 751-767.

- Langland-Hassan, Peter. 2011. "A Puzzle about Visualization." *Phenomenology and the Cognitive Sciences* 10 (2): 145-173.
- Levin, Michael. 2006. "Gettier Cases without False Lemmas?" *Erkenntnis* 64 (3): 381-392.
- Lewis, David. 1996. "Elusive Knowledge." *Australasian Journal of Philosophy* 74 (4): 549-567.
- Lu, Yuhao, Tom Stafford, and Charles Fox. 2016. "Maximum Saliency Bias in Binocular Fusion." *Connection Science* 28 (3): 258-269.
- McGrath, Matthew. 2017. "Knowing what Things Look Like." *Philosophical Review* 126 (1): 1-41.
- Michaelian, Kourken. 2016. "Against Discontinuism: Mental Time Travel and our Knowledge of Past and Future Events." In *Seeing the Future: Theoretical Perspectives on Future-Oriented Mental Time Travel*, edited by Kourken Michaelian, Stanley B. Klein and Karl K. Szpunar. Oxford: Oxford University Press.
- Mirza, M. B., R. A. Adams, C. Mathys, and K. J. Friston. 2018. "Human Visual Exploration Reduces Uncertainty about the Sensed World." *PLoS One* 13 (1): e0190429.
- Molyneux, Bernard. 2007. "Primeness, Internalism and Explanatory Generality." *Philosophical Studies* 135 (2): 255-277.
- Morrison, John. 2016. "Perceptual Confidence." *Analytic Philosophy* 57 (1): 15-48.
- Nagel, Jennifer. 2010. "Epistemic Anxiety and Adaptive Invariantism." *Philosophical Perspectives* 24: 407-435.
- . 2013. "Knowledge as a Mental State." *Oxford Studies in Epistemology* 4: 275-310.
- . 2017. "Factive and Nonfactive Mental State Attribution." *Mind & Language* 32 (5): 525-544.
- . 2021. "Losing Knowledge by Thinking about Thinking." In *Reasons, Justification and Defeat*, edited by Jessica Brown and Mona Simion. Oxford: Oxford University Press.
- Palmer, Colin J., Bryan Paton, Melissa Kirkovski, Peter G. Enticott, and Jakob Hohwy. 2015. "Context Sensitivity in Action Decreases Along the Autism Spectrum: A Predictive Processing Perspective." *Proceedings of the Royal Society B: Biological Sciences* 282 (1802): 20141557.
- Parr, Thomas, Andrew W. Corcoran, Karl J. Friston, and Jakob Hohwy. 2019. "Perceptual Awareness and Active Inference." *Neuroscience of Consciousness* 2019 (1): niz012.
- Parr, Thomas and Karl J. Friston. 2017. "The Active Construction of the Visual World." *Neuropsychologia* 104: 92-101.
- . 2018. "The Anatomy of Inference: Generative Models and Brain Structure." *Frontiers in Computational Neuroscience* 12: 90.

- Paton, Bryan, Josh Skewes, Chris Frith, and Jakob Hohwy. 2013. "Skull-Bound Perception and Precision Optimization through Culture." *Behavioral and Brain Sciences* 36 (3): 222.
- Peacocke, Christopher. 1979. "Deviant Causal Chains." *Midwest Studies in Philosophy* 4: 123-155.
- Pernu, Tuomas K. 2009. "Is Knowledge a Natural Kind?" *Philosophical Studies* 142 (3): 371-386.
- Pezzulo, Giovanni and Stefano Nolfi. 2019. "Making the Environment an Informative Place: A Conceptual Analysis of Epistemic Policies and Sensorimotor Coordination." *Entropy* 21 (4): 350.
- Plantinga, Alvin. 1993. *Warrant: The Current Debate*. New York: Oxford University Press.
- Pryor, James. 2000. "The Skeptic and the Dogmatist." *Noûs* 34 (4): 517-549.
- Quine, W. V. O. 1969. "Epistemology Naturalized." In *Ontological Relativity and Other Essays*, 69-90. New York: Columbia University Press.
- Rao, Rajesh PN and Dana H. Ballard. 1999. "Predictive Coding in the Visual Cortex: A Functional Interpretation of some Extra-Classical Receptive-Field Effects." *Nature Neuroscience* 2 (1): 79-87.
- Schiffer, Stephen. 2009. "Evidence = Knowledge: Williamson's Solution to Skepticism." In *Williamson on Knowledge*, edited by Patrick Greenough and Duncan Pritchard, 183-202. Oxford: Oxford University Press.
- Seth, Anil K. 2013. "Interoceptive Inference, Emotion, and the Embodied Self." *Trends in Cognitive Sciences* 17 (11): 565-573.
- Soteriou, Matthew. 2016. *Disjunctivism*. Abingdon: Routledge.
- Stine, Gail C. 1976. "Skepticism, Relevant Alternatives, and Deductive Closure." *Philosophical Studies* 29 (4): 249-261.
- Swanson, Link R. 2016. "The Predictive Processing Paradigm has Roots in Kant." *Frontiers in Systems Neuroscience* 10: 79.
- Weisberg, Jonathan. 2020. "Belief in Psyontology." *Philosophers' Imprint* 20 (11).
- Westra, Evan. 2019. "Stereotypes, Theory of Mind, and the Action–prediction Hierarchy." *Synthese* 196 (7): 2821-2846.
- White, Roger. 2014. "What is My Evidence that here is a Hand?" In *Scepticism and Perceptual Justification*, edited by Dylan Dodd and Ella Zardini, 298-321. Oxford: Oxford University Press.
- Wiese, Wanja. 2017. "What are the Contents of Representations in Predictive Processing?" *Phenomenology and the Cognitive Sciences* 16 (4): 715-736.

- Wilkinson, Sam. 2014. "Accounting for the Phenomenology and Varieties of Auditory Verbal Hallucination within a Predictive Processing Framework." *Consciousness and Cognition* 30: 142-155.
- Williams, Daniel. 2018. "Predictive Processing and the Representation Wars." *Minds and Machines* 28 (1): 141-172.
- Williamson, Timothy. 2000. *Knowledge and its Limits*. Oxford: Oxford University Press.
- . 2005. "Knowledge, Context, and the Agent's Point of View." In *Contextualism in Philosophy: Knowledge, Meaning, and Truth*, edited by Gerhard Preyer and Georg Peter, 91-114. Oxford: Oxford University Press.