Un-debunking ordinary objects with the help of predictive processing

Paweł Gładziejewski

Abstract: Debunking arguments aim to undermine common sense beliefs by showing that they are not explanatorily or causally linked to the entities they are purportedly about. Rarely are facts about the etiology of common sense beliefs invoked for the opposite aim, that is, to support the reality of entities that furnish our manifest image of the world. Here I undertake this sort of un-debunking project. My focus is on the metaphysics of ordinary physical objects. I use the view of perception as approximate Bayesian inference to show how representations of ordinary objects can be extracted from sensory input in a rational and truth-tracking manner. Drawing an analogy between perception construed as Bayesian hypothesis testing and scientific inquiry, I sketch out how some of the intuitions that traditionally inspired arguments for scientific realism also find application with regards to proverbial tables and chairs.

1   Introduction

Debunking arguments aim to establish that the mechanisms and processes that generate our manifest image of the world are systematically skewed by factors unrelated to truth. Specifically, the debunkers argue that sometimes what we represent the world to be like is not properly explained by the world actually being that way. This way, the facts about the genealogy of common sense beliefs (experiences, intuitions, and so on) are said to undermine or defeat their justification (or the ability to confer justification).

The debunking strategy has been employed to undermine the justificatory status of beliefs about a number of domains, including moral values, God(s), colours, selves, the passage of time or ordinary objects. The debunkers aim establish that those beliefs are not caused by whatever they purport to be about. Rather, they are produced by contingent social and cultural factors, or by evolutionary causes related to reproductive success, or by cognitive mechanisms geared towards representing the world in terms of useful fictions rather than its real structure (see Korman, 2019 for an extensive review of debunking arguments).

Notice that although etiological considerations are commonly used for revisionary purposes, they are not *a priori* bound to lead to debunking conclusions. In principle, knowing the etiology of a belief (experience, intuition, and so on) could increase its justificatory status and hence 'un-debunk' the relevant domain. However, it seems that this possibility is rarely seriously considered in the literature.

In this paper, I aim to provide a proof of concept regarding how a naturalistic account of origins of certain intentional states could support realism about the domain that these states are about. At the very least, I hope to show that we always need to get the details of the etiological story exactly right to rationally evaluate whether a belief is successfully debunked or not.

My focus will be on the metaphysical status of ordinary physical objects, insofar as our assessment of this status is guided by knowledge about cognitive processes underlying object perception. These are the processes that structure raw data streams registered by the senses into perceptual representations of unified objects.

I will start (in Section 2) by discussing existing attempts at debunking ordinary objects which point to the purported fact that the processes behind object perception are disconnected from how the world is really structured (Goldman [1987], [1993]; Osborne [2016]). I will try to show that these arguments are inconclusive and leave the question about the reality of ordinary objects wide open.

Next, I will lay out my positive case. I will start by introducing (in Section 3) the predictive processing view of object perception. On this story, perceptual representations of objects are an output of probabilistic inference, where the inference is realized through minimizing the prediction error. I will argue (in Section 4) that predictive processing supports my un-debunking project in two ways. First, on such a view, the transition from sensory input to representations of objects is epistemically rational. Second, effectively minimizing the prediction error requires the cognitive system to construct an accurate (to some approximation) model of the causal structure that produces the sensory input. Hence, the process behind object perception is not only rational, but also truth-conductive.

To further support my case, I will (in Section 5) follow a tradition of thinking about perception by analogy to scientific inquiry (Gregory [1980]). The goal will be to point out that some intuitions that traditionally drove arguments in favor of scientific realism can also be put to use to support realism about ordinary objects. Drawing from philosophy of science, I will sketch out perception-focused variants of the no-miracles argument and the argument from convergence to the truth.

I will conclude by briefly discussing some further issues and adding some caveats regarding the position defended here.

2 Debunking ordinary objects with the cognitive science of perception?

Cognitive-science-based attempts at debunking ordinary objects focus on mechanisms of perception. In particular, they focus on the processes by which the cognitive system structures the raw manifold of sensory signals into perceptual representations of stable units like cats, chairs, or pizzas (Goldman [1987], [1993]; Osborne [2016]). For this strategy to be fully successful, one presumably needs to assume that because our beliefs about which objects exist ultimately originate from perceptual experience, debunking object perception automatically debunks corresponding beliefs. I will follow this perception-centered approach here, although it is not entirely unproblematic (one might argue that there is a reciprocal connection between perception and beliefs, whereby beliefs also affect perception; I will come back to this in Section 4).

So, take a case where, by virtue of the workings of your visual system, the retinal input gives rise to a perceptual representation of a visual scene composed of ordinary objects. To use a common example (Korman [2015]), assume that your perception reveals to you a scene involving a dog and a tree. But we can imagine a visual system which interprets the same input by binding the dog and the tree trunk to compose a single 'trog'. And there may be still other, even more ontologically exotic (for humans) ways of interpreting the same input which defy description in natural language.

The question is whether the structure of the world explains why we see trees and dogs rather than trogs or other strange entities. Intuitively it may seem that making sense of the sensory input in terms of ordinary objects (but not strange objects) gets thing right. It cuts the world at

its joints. Perhaps there really are dogs and trees, which straightforwardly explains why we experience the world the way we do.

The core of debunking the common sense metaphysics of objects lies in denying this intuitive explanation. The debunkers set out to provide an alternative explanation. Specifically, the sort of debunking strategy I'm interested in here aims to establish that what really explains which kinds of objects we take to exist are the rules, principles, or algorithms which our cognitive/perceptual mechanisms use to make sense of sensory input. And knowing those rules, principles, or algorithms supposedly reveals that perception is systematically off-track, that is, it does not track the truth about which objects exist (Osborne [2016]). This in turn defeats whatever justification we may have had for believing in ordinary objects.

Which facts about the processes behind object perception are invoked to support the debunking project? Goldman ([1987]) points to the famous Gestalt principles of perceptual grouping, by which elements of sensory manifold get grouped together to yield experiences of unified objects or 'figures'. These include, for example, the principles of proximity, similarity, or good continuation. In later work, Goldman ([1993]) further develops his case by invoking Spelke's (see Spelke [1990]) account of principles of object perception. He points to evidence for the claim that humans have an innate, early-developing mechanism for object perception which tracks units that exhibit cohesion, are clearly bounded, or move in a spatiotemporally continuous manner.

More recently, Osborne ([2016]) has extended the project of providing 'off-track' accounts of object perception by pointing out some additional facts about the workings of the visual system. First, he notes that any two-dimensional retinal input image underdetermines the interpretation in three-dimensional terms, in that it is consistent with a number of (in principle infinite) such interpretations. To work around this problem, and to compensate for inherent processing power limitations, the brain makes use of computationally cheap heuristic short-

cuts. Second, visual object recognition is context-dependent and based on functional relations. For example, Osborne ([2016]) mentions studies showing that contextual and functional factors affect perceptual categorization of ambiguous figures.

The lesson we should presumably draw from all those considerations is as follows. Rather than being led by the metaphysical truth itself, our perceptual experience of objects is based on heuristics, Gestalt principles and factors related to cohesiveness, continuity, contextual, or to functional relationships. The assumption is that those principles and factors are irrelevant with respect to the issue of which objects really exist.

Admittedly, Goldman and Osborne are right to notice that the accounts of processes underlying object perception do not need to directly invoke facts about the metaphysical structure of the world. And yet, their arguments are far from conclusive. What is crucially missing is a positive case for thinking that the principles of object perception that these authors cite are in fact off-track with respect to metaphysical truth. It seems that for the debunking project to work, one would need to presuppose that the truth about what objects exist is such that the principles underlying object perception are ill-suited to reveal it, or that reality lacks any sort of unity or objecthood principle that corresponds to what we experience (Schaffer [2019] makes a similar point about Goldman's argument). But nothing said so far guarantees that this presupposition is right. Perhaps to get to an on-track explanation of object perception, we simply need to look further down the explanatory chain and ask why our perceptual and cognitive mechanisms structure the sensory input according to this set of principles rather than some other set. It is still a wide open possibility that that the correct answer to this question will have to do with the truth-conductive nature of this particular set of principles.

Furthermore, there are reasons to doubt just how strongly some of the facts invoked by Goldman and Osborne support the debunking project. The Gestalt principles, for example, are unlikely to be just arbitrary ways of structuring the sensory input. It turns out that the these

principles capture statistical patterns that recur in natural scenes, for example, the geometry of sequences of edges (see Elder and Goldberg [2002]; Geisler [2008]; Sigman *et al*. [2001]). This suggests that the Gestalt principles are 'ecologically valid' or tuned to order present in the environment that produces visual images (which of course, by itself, does not yet imply that this order is connected to facts about objecthood). A similar point can be made about the use of heuristics by the cognitive system. It has been forcefully argued that heuristics that people use to solve problems usually yield accurate results because they fit the statistics of the environment (Gigerenzer and Brighton [2009]). Arguably, this may well apply to the heuristics that the visual system relies on to make sense of the retinal input. So the tension between accuracy and efficiency of processing may be merely apparent, and the visual system does not necessarily have to trade one for the other.

However, a deeper diagnosis can be made which I think reveals the crux of where exactly the debunking attempts just discussed go wrong. The crucial theme running through Goldman's and Osborne's arguments is the recognition of just how 'constructive' or 'interpretative' perception is. What these authors point to is the fact that what we experience is a heavily processed end-result of computations to which we have no conscious access. Perhaps this idea could undermine a naïve or direct realist account of object perception, grounded in (some versions of) neo-Gibsonian or enactive approaches.[1] But there are other ways to be a realist about objects than being a direct realist. The simple transition from the constructive nature of perception to the idea of it being systematically off-track with respect to the structure of the world would constitute a manifest *non-sequitur*. Just because a representation is generated by a sophisticated computational process does not in any way entail or even indicate that it is inaccurate, or likely to be inaccurate. The choice, then, is not between a story on which our experiences are immediately glued to the structure of reality and one on which they are actively

---

[1] I thank an anonymous reviewer for pointing this out.

constructed. Instead, even if we agree on a constructivist view of perception, the issue is whether all the information processing involved in perceiving objects is such that it will (tend to) produce accurate representations of what entities are 'out there'.

This way, we may distinguish two alternative ways of developing a constructivist account of object perception, each of which has drastically different bearing on whether the debunking project succeeds.

On a debunking-friendly version of the constructivist view, using perception to navigate the world could be regarded as analogous to using a computer desktop interface (Hoffman *et al.* [2015]). As I move a red icon on my desktop toward a trash icon, I succeed in deleting a file. But this does not entail that there really is a red object in the computer that is literally being moved around until it disappears. Ordinary objects, as revealed in perception, could be like icons on a computer screen. Perceiving the world as composed of cats and rocks could subserve adaptive action, even if cats and rocks do not correspond to any real, mind-independent worldly structures. On this approach, then, perceptual representations of objects would constitute a form of adaptive misbeliefs, that is, beliefs that are false 'by design', but which nonetheless enable successful action (McKay and Dennett [2009]; see also Akins [1996]).

Consider now a different analogy, which suggests an alternative, realist-constructivist view. Cartographic maps are action-guiding devices, but they afford pragmatic success only under the condition that they accurately (if approximately or in an idealized way) represent some referent spatial structure. Perhaps, then, navigating the world using perception is more akin to employing a cartographic map to find one's way around the environment (see Gładziejewski [2016]). The constructivist view of perception as such is perfectly consistent with the possibility that perceptually guided action requires constructing a largely accurate representation of the environment; it is also consistent with the possibility that the processes which turn raw sensory data into perceptual representation proceed in accordance to rational rules.

It is this latter, realist view that we should subscribe to, I will claim, if perceptual representations of objects are constructed in accordance with the Predictive Processing story.

3   Predictive processing: object perception as causal inference

3.1 A brief overview of predictive processing

Predictive processing (henceforth PP) belongs to a larger historical lineage of understanding perception as unconscious inference (Gregory [1980]; Helmholtz [1867]).[2] Underlying this view is the recognition that sensory signals are ambiguous. There is no one-to-one mapping from worldly causes to sensory input, and the input itself is inherently noisy. The job of perception thus lies in inferring the most likely causes of sensory stimulation. In PP, this inference is taken to approximately conform to the Bayes rule:

$$P(H|E) = P(E|H) \, P(H) \, / \, P(E)$$

In essence, the Bayes rule is a recipe for selecting a hypothesis most likely to be accurate, given some evidence (for example, the sensory signal). This is the posterior probability of the hypothesis, *P(H|E)*. To compute the posterior, a Bayesian subject needs to consider both the prior belief about the probability of the hypothesis being true, *P(H)*, and the likelihood of the hypothesis, which expresses the belief about how likely it is obtain the evidence, assuming the truth of the hypothesis, that is, *P(E|H)*.

---

[2] For sake of brevity, I only describe core tenets of PP and highlight those aspects of the theory which are directly relevant for later discussion. For much more detailed and comprehensive treatments, see, for example. (Clark [2016]; Friston [2005]; Hohwy [2013]).

Bayesian inference can be cast as a process of minimizing prediction error. Assume that the prior is a Gaussian probability distribution over samples of evidence. Assume further that the samples actually obtained over some period also conform to a Gaussian distribution. The distance between the means of both distributions is the prediction error. A posterior estimate is formed that updates the prior in light of the prediction error. In the 'empirical Bayes' framework employed in PP, the posterior turns into a prior in the next iteration of inference, and then a new posterior is formed in light of consecutive prediction error. Importantly, the degree to which the posterior is determined by the history of learning (the prior) versus the need to fit data (the likelihood) depends on the relative precisions (inverse variances) of respective distributions. This is because precision is indicative of their reliability for belief updating.

Now, on the PP view, the brain stores a predictive, generative statistical model of the environment. The model is 'generative' because it encodes information about how the sensory input depends on the causal structure that generates it. Specifically, the model captures the causes of the sensory input as well as the way they dynamically interact with each other.

The generative model is hierarchically structured, which reflects the nested nature of the causal structure in question. Lower levels predict the flow of rapidly changing regularities in sensory input, while higher levels predict more invariant patterns which emerge over longer time-scales.

Crucially, the generative model is constantly optimized to minimize the (average) prediction error. The error is encoded by the bottom-up signals which indicate the discrepancy between the predictions generated at a given level of the model and the actual states of the level adjacently below. This process ultimately bottoms out in the level that predicts the sensory signal itself. Across the hierarchy, the error signals are weighted according to their estimated precisions, so the relative contribution to processing of top-down predictions and bottom-up error signals is flexibly regulated. In PP, estimating precisions accounts for attention.

On the PP picture, it is through its reliance on minimizing the prediction error—rather than by explicitly computing posterior distributions—that perception gets to count as approximately Bayesian (Hohwy [2013]).

The description so far focused on perceptual inference, where the brain is trying to fit its estimates to sensory signals. But PP is also (and arguably primarily) put forward as an explanation of action. Here, action is 'active' inference. Roughly, action is intervening on the causal structure to induce streams of sensory evidence that conform (or not) to some prior hypothesis. If successful, action minimizes the prediction error. Of particular interest for me here is epistemic active inference, that is, actions aimed at reducing uncertainty about the causal structure producing the sensory signal (see Clark ([2018]) for a philosophical discussion of the 'pragmatic' and epistemic active inference, where the distinction is cast in terms of exploration/exploitation dilemma). For example, saccadic movements can be construed as 'micro-experiments' which selectively sample sensory evidence to discriminate between two competing perceptual hypotheses about the causes of input (Friston *et al.* [2012]).

On the PP view, learning the generative model is, so to speak, supervised by the environment (Hohwy [2013]). Ultimately, what the cognitive system tests its model against are the statistical patterns that arise at its sensorium. And the latter depend on the structure of the external causal milieu itself. Computational work employing hierarchical Bayesian models shows that even 'abstract' priors can be induced from raw, unlabeled data, without relying on stores of pre-existing (for example, innate) knowledge (cf. Kemp *et al.* [2007]; Perfors *et al.* [2011]). In this sense, PP-based systems exhibit the capacity to construct their models by relying exclusively on data accessible from the 'animal's perspective' (Eliasmith [2005]; see also Clark [2013]).

3.2 Ordinary objects as common causes

The question of how perceptual representations of stable, unified objects are generated can be naturally cast in terms of feature binding. There are empirical results that strongly suggest that perceptual features are represented before they become integrated into representations of objects (Treisman and Schmidt [1982]). These features are only subsequently bound together as belonging to unified entities. For example, in vision, there is presumably a stage of processing at which an unorganized set of colours, shapes, and textures is turned into a representation of a scene composed of objects that have those features. Somehow, the mechanisms of perception 'decide' which features properly go together and should be bound, and which do not.

According to PP, feature biding is a form of inference, where a common cause is inferred as responsible for a pattern of features (see Hohwy [2013], Chapter 5; Wiese [2018]; for similar proposals, see Körding *et al.* [2007]; Parise *et al.* [2012]; Shams and Beierholm [2010]; Witkin and Tenenbaum [1983]). Patterns are predictable regularities. The job of the generative model is to learn to predict those regularities by estimating their underlying causes. With prediction in place, all the information required to update the estimates can be compressed to a prediction error signal. Crucially, a sensory pattern can sometimes be accounted for (and predicted) in terms of a common cause that produces and sustains it. In fact, because in realistic scenarios the input is generated by multiple interacting causes, the sensory signal (for example, the retinal image) is treated as a mixture of signals coming from multiple sources. So the generative model needs to 'unmix' the signal to estimate the number of distinct causes, and infer which features 'belong' to which common cause. And those inferred common causes are ordinary objects, as represented in perception.

Consider a very simplistic case where the generative model produces a low-level representation of features. In the visual noise, a smooth sequence of roughly colinear edges is

discerned. The edges close in to form a shape. The shape does not disperse but is retained over time. There is a patch roughly uniform in colour and texture. The position of the patch and the way it changes over time overlaps with the movement of the shape. Thus, the shape, colour, and texture are correlated to form a predictable regularity. The regularity is unlikely to be accidental. At a higher level of the generative model, this pattern is used to infer a single underlying cause, which in turn drives predictions about the unfolding of processing at the level below. This way, the environment is represented in terms of a single object that bounds the lower-level features.[3] Under a common cause explanation, it can be predicted that (1) manipulating the object should result in correlated change of the bound features (for example, the features will remain spatiotemporally continuous as the object is moved from one place to another); (2) the features themselves are mutually statistically independent, conditioned on the common cause (given the position of the object, the position of the colour patch provides no additional information about the position of the shape; and it should be possible to change one of the features while keeping the other one constant). Those predictions can be put to test in epistemic active inference. If the latter induces sensory streams that conform to the predictions, the common cause hypothesis gains evidence in its favor.

Interestingly, over the last two decades, much experimental and modelling work on the role of causal inference in perception has focused on multisensory integration. Here, the integration of signals arising from multiple sensory channels (for example, a sight of a moving winged object and hearing a moving chirping sound) is construed in terms of inferring a single cause (a bird) that accounts for the correlations between those signals. Three findings are worth

---

[3] PP does not have to be committed to a view that objects are represented as bundles of properties. In principle, once the features have been bound, the resulting object could be tracked in terms of regularities which abstract from the features themselves (for example, spatiotemporal continuity), so that it could be represented as retaining identity despite complete change of features.

mentioning here. First, multisensory correlations on which the brain relies when inferring common causes are related to similarity of temporal structures between modality-specific signals (see, for example, Parise *et al*. [2012]). Second, the multisensory estimates are affected by how reliable the respective modalities are (see, for example, Ernst and Banks [2002]). Third, multisensory integration can be fruitfully modeled as Bayesian model selection, where the correlations between sensory signals are used to decide between two competing causal models of how the signals were generated (see, for example, Körding *et al.* [2007]).

3.3 Predictive processing, other explanations and 'strange' objects

Two things are worth pointing out before I move on. First, the PP view does not have to be treated as an alternative to the explanations of object perception discussed by Goldman and Osborne. Instead, those accounts are not only largely consistent, but PP can be plausibly seen as revealing a deeper underlying rationale for the principles invoked by Goldman and Osborne. If Gestalt principles describe regularities in retinal images produced by natural scenes, it is to be expected that PP-based cognitive systems will rely on those regularities (by predicting them) to make sense of sensory input. A similar story could be told about 'Spelke-principles', in that boundedness, cohesion, or spatiotemporal continuity can by hypothesized to constitute reliable patterns that brains can rely on when predicting the flow of sensory states. In his paper, Osborne ([2016]) discusses the 'generic viewpoint' heuristic, whereby the visual system assumes that the scene perceived is not accidentally produced by a special viewpoint of the observer. However, this heuristic can be naturally treated as a Bayesian prior that, on average, contributes to maximizing posterior probabilities of visual estimates (Freeman [1996]). Thus, arguably, a substantial part of the material cited in support of debunking ordinary objects can be subsumed under the present un-debunking project.

Second, an important feature of PP is that it does not rule out the possibility that our cognitive and perceptual mechanisms could represent the environment in terms of objects alien to our common sense. Furthermore, PP suggests that there are principled reasons as to why humans do not carve up the world in accordance to strange ontologies. Think of different perceptual ontologies as competing causal models of how the sensory input is produced. In PP, what arbitrates between those models is how well they deal with the job of minimizing the (average) prediction error. Arguably, models invoking 'strange' objects should be ruled out because they are ineffective at minimizing the prediction error, which in turn stems from the fact that they get the causal structure which generates the sensory input wrong.

Take a trog, an object composed of a dog and a tree trunk. On the present view, a trog is a perceptual hypothesis or estimate that attributes the 'doggy' and 'tree-trunky' part of the retinal image to a common worldly cause. PP does not preclude the possibility of a visual system coming up with such hypothesis. But PP also shows why it should reject it. Consider a simpler case of seeing a cat behind a fence. As long as a static stimulus is considered, it could be explained by invoking slices of a cat interposed between parts of the fence. But as one moves relative to the scene, the unfolding of visual input reveals previously unseen parts of the cat, which straightforwardly favors the interpretation in terms of a (whole) cat behind the fence (Hohwy [2013]).

By analogy, the 'doggy' and 'tree-trunky' parts of the visual input do not participate in a stable pattern that would merit a common cause explanation. Put simply, they are not correlated enough; one does not predict the other reliably enough for them to count as having a common origin. We should expect a model which treats them as a single object to prove relatively ineffective at minimizing the prediction error. For example, a bit of epistemic active inference should reveal that the doggy-part of the visual stream is only weakly (if at all) predictive of the

tree-trunky part and vice versa. Overall, a model that treats dogs and a tree trunks as distinct entities should fare better at predicting the flow of sensory input.

PP also provides a principled explanation as to why we do not perceive the world in terms of non-detached-arbitrary-parts of ordinary objects. Consider a causal model which parcels a given subset of visual input into a number of non-detached-arbitrary-rabbit-parts (henceforth NDARPs) rather than simply attributing it to a rabbit. A debunker of ordinary objects might stipulate that models which invoke rabbits and NDARPs are predictively equivalent and hence will do equally well at minimizing the prediction error.

However, there are strong reasons to think that predictive brains should posit rabbits rather than NDARPs. It is well-established that Bayesian model selection should exhibit preference for simpler models that invoke fewer factors to explain data (Feldman [2016]). This is especially so if the increased complexity is not compensated for by better fit to data. On those grounds, it has been argued that PP-based cognitive systems employ mechanisms that aim to reduce the complexity of their generative models (Friston and Hobson [2012]).[4] Now, a generative model which postulates NDARPs treats a part of the retinal input as produced by a number of underlying causes, each corresponding to a distinct arbitrary rabbit part. A model which simply invokes rabbits is simpler, as it assumes that the same part of the image has a single cause. Even under the provisional assumption that both models can be predictively on par (which I will criticize below), a model invoking NDARPs does not get any leverage to compensate for its complexity. Thus, on grounds of parsimony, a Bayesian cognitive system should select a model which represents the environment in terms rabbits.

---

[4] Independently of PP, it has also been suggested that the preference for simplicity in probabilistic inference may explain the Gestalt rule Prägnanz, which (roughly) consists in favoring parsimonious groupings of the sensory input (Feldman [2016]).

Furthermore, considerations regarding simplicity also cast doubt about whether we should grant the very assumption that rabbits and NDARPs can be equally predictively effective. After all, parsimony is preferable in model selection precisely because it is positively related to predictive accuracy. More complex statistical models tend to suffer from overfitting the noise, and consequently turn out less predictively accurate (for an extensive philosophical discussion, see Forster and Sober [1994]). In the long run and on average, then, we should expect the simpler rabbit model to produce less prediction error than the more complex NDARPs model.

Lastly, notice that for a collection of NDARPs to predict the sensory input as well as the rabbit model, they would need to be bounded at a higher level of the generative model.[5] However, this would effectively mean forming a rabbit hypothesis to explain sensory input. So to achieve predictive equivalence, the NDARPs ontology would have to collapse, at some level of the generative model, into a rabbit ontology. Also, by definition, we are talking about arbitrary composite parts which may comprise, say, a single 'object' composed of the rabbit's ear, its right eye and the tip of its tail. But if so, then the points raised against trogs also apply to NDARPs. Just as predictive perceptual systems should favor dogs and trees to trogs, they should prefer to parcel the rabbit into non-arbitrary composite parts, that is, parts which are themselves properly interpretable as objects (like the rabbit's ear or its tail).

4 Ordinary objects un-debunked

The picture laid out in the preceding section provides us with an account of the process which connects, *via* sensory input, whatever it is 'out there' to perceptual representations of the world as populated with ordinary objects. I claim that this constitutes an on-track explanation of object perception. It renders the process of object perception appropriately responsive to the

---

[5] I am indebted to an anonymous reviewer for suggesting this line of reasoning.

actual structure of the environment. Therefore, in light of PP, the epistemic status of our beliefs in dogs, rabbits, or pizzas is left largely unscathed.

Two features of the PP account render perception on-track. First, PP proposes that the process leading from sensory stimulation to perceptual representations of objects is epistemically rational. Second, it postulates that minimizing the prediction error in perception requires building a true or accurate model of the causal structure that produces the sensory input. I discuss these two features in turn.

On the PP account, the rationality of the transition from raw sensory data to perceptual representations of objects lies in its approximate conformity to a rational rule of inference.[6] The rule in question is, of course, the Bayes rule. The process of updating perceptual representations is such that it will tend to maximize the posterior probability of those representations, construed as hypotheses or estimates. This way, to make sense of current sensory input, the brain will tend to select representations most likely to be accurate, given this input (and prior beliefs). We see dogs and trees rather than trogs or collections of arbitrary parts of objects because hypotheses invoking the former are attributed higher posteriors than the hypotheses invoking the latter.

Furthermore, the process of updating perceptual representations marries two principles: the need to remain coherent with one's history of learning (priors) and the need to fit current sensory evidence. In PP, the relative influence of priors and the need to fit data is itself rationally regulated through estimating the relative precisions (hence, reliability) of priors and the sensory evidence.

Given the rationality of the process underlying object perception, it is no longer clear how knowing about this process could defeat beliefs about ordinary objects. On the contrary, it is

---

[6] Space forbids me from defending the view that the sub-personal (non-conscious, non-voluntary) processes of the type that PP posits can count as rationally evaluable inferences. See (Kiefer [2017]) for an detailed defense of the view that the notion of inference at use in PP should be taken literally.

natural to think that the rationality of the perceptual process confers a positive epistemic status on perceptual representations of objects (see Siegel [2017] for an influential defense of the view that the rational etiology of a perceptual state affects the epistemic standing of this state). This positive status could be transferred to the beliefs about objects, especially if the move from perception to belief is itself Bayesian (see Brössel [2017]).

What I have said so far is purposefully noncommittal with respect to the exact architecture of perceptual justification that falls out of PP (for some proposals, see Ghijsen [2018]; Gładziejewski [2017]; Munton [2018]). For present purposes, it suffices to establish that perceptual representations of objects are rationally produced. One potential worry requires discussion, though. On the PP view, perceptual estimates are driven both by sensory signals and the prior beliefs encoded in the generative model. Hence, the rational standing of perceptual representations is in part conditional on the rational standing of the generative model. But the priors encoded in the model may themselves be catastrophically off-track. If this was the case, it would undermine the epistemic status of the perceptual hypotheses.

I think that PP has the theoretical resources to deal with this issue. The crucial observation is that the brain, on the PP view, is not a dogmatic system. Over time, the priors themselves can be reshaped in light of the prediction error, a process which ultimately bottoms out with sensory stimulation itself. Thus, the priors across the hierarchy are responsive to sensory evidence. Consider priors stored at high levels of the generative model. They encode sensory patterns that unfold over long time-spans. For example, rather than encoding information about ordinary objects (cats, trees, desks, and so on), they may work by constraining the patterns through which objects are tracked at a lower level of the model ('Track entities that are cohesive, spatiotemporally continuous…'). But, as mentioned in the previous section, under hierarchical Bayesian computational schemes that PP trades in, these high-level estimates may themselves be induced from sensory data. More generally, PP postulates that perceptual updating and

perceptual learning (acquiring and revising priors) are both explained in terms of minimizing the prediction error. The difference lies in the temporal grain of the process, not in its nature. Furthermore, even if we assumed that many priors are both innate and initially off-track, this would not necessarily undermine the present point. It has been empirically established that when exposed to contrary sensory evidence, the visual system can revise even the priors traditionally thought innate or hard-wired, like the assumption that light comes from above or that objects are either stationary or move at slow speeds (cf. Adams *et al*. [2004]; Sotiropoulos *et al*. [2011]).

A critic might point out that the discussion so far does not in any way establish that our experiences of objects track mind-independent reality. Perhaps perceptual experiences of objects enjoy a positive epistemic status while being systematically false. The folk may be rationally entitled to believe in tables and chairs or be epistemically blameless for doing so. But they are still mistaken. To address this issue, I want to discuss the second feature of PP that makes it an on-track sort of explanation. The crux is this: cognitive systems can manage to minimize the prediction error only if they accurately represent the causal structure generating the sensory input.[7] This not only establishes PP as an on-track explanation of object perception,

---

[7] Note that in the present paper, I investigate the debunking project in terms of the relation between the common sense ontology of objects and the actual causal structure that generates the sensory data. However, there is an influential, mereologically-oriented debunking argument applied to ordinary objects that I do not address here at length. Roughly, this argument centers around the observation that even if facts about mereological composition of reality were such that ordinary objects do not exist (for example, there are no chairs, only collections of atoms arranged chairwise), perceptual mechanisms would still produce the experience of ordinary objects (see Korman [2015] for extensive discussion). Two things may be said about this mereology-oriented debunking argument in the present context. Frist, the present treatment is not meant to offer a definitive answer to it. If there only existed atoms-arranged-chairwise, they would still generate sensory patterns that would cause predictive mechanisms to produce perceptions as of chairs. This should not be seen as surprising, as the mereological debunking argument

but does so without begging the question by simply presupposing what sorts of objects exist, and then postulating that they are explanatorily linked to the content of perceptual experience.

To understand the present point, we need to focus on the notion of representation in use in PP (I will touch on non-representational approaches to PP in the concluding section). A number of philosophers working on PP have converged on a view that the generative model acts as a cognitive representation of the environment (Gładziejewski [2016]; Kiefer and Hohwy [2018]; Williams [2017]). In particular, it constitutes a representation grounded in exploitable structural similarity. This essentially means that: (1) the generative model encodes a hierarchical relational structure of latent or hidden variables; (2) based on this structure, the model produces sensory predictions, which in turn are tested against actual input, giving rise to prediction error; (3) the capacity to minimize the prediction error (also through action) depends on the degree to which the model's relational structure maps onto the nested causal structure which produces sensory states. For example, the conditional dependencies between latent variables may map onto the causal relations between respective entities in the world, thus contributing to successful prediction of the sensory input. More generally, the degree of structural similarity between the model and the causal dynamics of the environment determines the degree to which the former accurately represents the latter (see also Gładziejewski and Miłkowski [2017]; Kiefer and Hohwy [2018]; Lee [2018]).

---

does not depend on the nitty-gritty scientific details of how perceptual mechanisms work anyway, and is thus beyond the scope of this paper (addressing the argument would require, for example, taking a closer look at the relation between ordinary objects and their underlying physics). Second, I think that casting the discussion about debunking in terms of causal structures and the data they explain has the advantage of making the issue more scientifically tractable. By analogy, in philosophy of biology, it might be more fruitful to frame the discussion about the existence of genes in terms of whether 'gene-talk' tracks causal structure(s) that produce patterns of trait inheritance, rather than in terms of whether the inheritance patterns could have been produced by atoms-arranged-genewise.

The crucial observation is that if the model is learned or optimized through minimizing the prediction error, then its structure is molded by the structure that produces the sensory states, whatever that latter structure may be. This way, over time, the model will come to structurally recapitulate the causes of the input. This is why learning or optimizing the model—as long as it subserves successful minimization of the (long-term, average) prediction error—involves increasing its representational accuracy.

More technically, the degree of representational (in)accuracy can be expressed here in terms of the Kullback-Leibler divergence between the generative probability distribution encoded in the top-down predictions, and the recognition distribution encoded in sensory and prediction error signals (see Kiefer and Hohwy [2018] for a detailed account). The KL-divergence takes the log ratios of probabilities of sensory evidence (under the recognition and generative model, respectively) to measure how much uncertainty about the recognition distribution remains when it is approximated using the generative distribution. The point is that the lower the divergence, the less inaccurate the generative model (Kiefer and Hohwy [2018]).

The present proposal is that the common sense ontology of objects proves effective at minimizing the KL-divergence between the generative and recognition distributions. That is, there is a level of the causal nexus which generates sensory input that is most accurately represented in terms of common sense ontology of objects. By employing an ordinary object ontology, the brain increases the representational accuracy of its generative model. In particular, the divergence is lower on this model than on a model which furnishes the world with trogs, NDARPs, or other strange objects.

Crucially, all those considerations are not meant to suggest that when perceiving objects, humans entertain a pristine 'view from nowhere', untouched by limitations of their cognitive apparatus. On the contrary, the present proposal is consistent with the fact that perception reveals only a fraction of the objective world. Perception can only access patterns in nature that

are discernible through sensory organs humans come equipped with. And the latter are only responsive to certain physical energies, within certain ranges. Using this limited sensorium, we extract models of the causal structure that produces sensory states, but our models only dig as deep into this structure as it makes sense for creatures designed primarily for survival and reproduction. Perception gives us no epistemic access to the world at the microscopic or cosmological level. And even within the restricted domain of medium-sized entities, the models that we use to make sense of sensory data may be in many ways simplified or approximate.[8] Still, within those limitations, it is not up to perceivers to decide which patterns are there in the sensory stream. We can only predict the sensory flow as long as our generative models track real patterns (Dennett [1991]; Ladyman and Ross [2007]). The ontology involving dogs and trees fares better at this than a trog-based ontology. My claim here is that the real patterns to which have access reveal the world as populated with ordinary objects. This implies that object perception is accurate (barring illusions or occasional misperception), even if it is also selective in an organism-relative way. This is different from it being 'projective' or systematically inaccurate.

5   Ordinary objects as the brain's theoretical posits: a case for realism

Consider a following characterization of scientific realism proposed by Hacking :

Scientific realism says that the entities, states and processes described by correct theories really do exist. Protons, photons, fields of force, and black holes are as real as toe-nails, turbines, eddies in a stream, and volcanoes. The

---

[8] The value of the KL-divergence can only be $0 \leq$. As long as there is any prediction error, the divergence will be greater than 0, so our internal models must misrepresent to some degree (Kiefer and Hohwy, 2018).

weak interactions of small particle physics are as real as falling in love.

(Hacking [1983], p. 21)

Notice how this contrasts with an assumption that seems to tacitly accompany the attempts to debunk the posits of common sense. From the perspective of a philosopher interested in debunking the common sense by invoking a scientific account of the origins of folk beliefs, Hacking gets the relative epistemic standing of the manifest and scientific image backwards. Doubt is cast upon the reality of toe-nails, volcanoes, or love. That we are rationally entitled to believe in the posits of successful scientific theories—or that we are substantially more entitled to this than to believe in the posits of common sense—seems to be taken for granted.

But whether scientific realism is a well-grounded position is, of course, far from settled. I want to draw on this simple fact to suggest a way to render realism about ordinary objects more reasonable. Under the unconscious-inference view, perception is sometimes treated as non-trivially akin to scientific inquiry (Gregory [1980]). For example, ordinary objects can be construed as theoretical entities that the brain posits to explain the patterns in sensory signals. In PP, the analogy between science and perception is deepened further because epistemic active inference plays a role similar to experimental interventions (Hohwy [2013]).

I want to recommend the following position. Under PP, the rational standing of the belief in ordinary objects may be largely on par with the rational standing of scientific realism; or at least the gap between them may be smaller that it initially appears. A case can be made for the reality of ordinary objects *qua* theoretical posits that is similar to the case that can be made for the reality of posits of scientific theories. Hence, a believer in photons may be rationally compelled to believe in toe-nails or volcanoes. This is because, from the perspective of PP, some of the core intuitions that traditionally drove arguments in favor of scientific realism also

find application in the context of the question about the reality of ordinary objects. I will discuss two such intuitions.

The first intuition underpins the famous no-miracles argument (Putnam [1975]). It essentially comes down to a following conjecture: it would take a miraculous coincidence for our scientific theories to achieve the predictive success that they do without them being at least approximately true. Here, truth is seen as the best explanation of empirical success. Consider now the generative model, as understood in PP. It is a statistical model learned and revised to produce ever more accurate predictions about the sensory signals, which are in turn produced by worldly causes. How are we to account for the generative model managing to successfully quash the prediction error? Its being accurate with respect to the causal structure generating the sensory signals is a plausible explanation. Only through a miraculous statistical fluke could a generative model that fails to capture the referent causal structure still manage to keep the prediction error low for prolonged periods of time. In fact, in PP, the no-miracles intuition gets some genuine traction because a plausible story can be said as to why it is legitimate to move from the predictive success to truth or accuracy (this is less obvious in the context of science). As discussed in the previous section, under the structural-representation view of the generative model, the ability to minimize the prediction error depends on the degree of structural match between the model and causes of the sensory signal; and the latter determines the representational accuracy of the model.

The second intuition relates to the notion of convergence to the truth, famously expressed by Peirce:

Different minds may set out with the most antagonistic views, but the progress of investigation carries them by a force outside of themselves to one and the same conclusion. This activity of thought by which we are carried,

not where we wish, but to a fore-ordained goal, is like the operation of destiny.

No modification of the point of view taken, no selection of other facts for study, no natural bent of mind even, can enable a man to escape the predestinate opinion. This great law is embodied in the conception of truth and reality. (Peirce [1878/2011], p. 63)

Peirce captures here an intuitively realist picture of scientific progress. As we apply the scientific method, new data is discovered and we revise our theories in light of it. Over time, false theories are discarded, errors are corrected, the disagreement between the scientists dissipates and the science converges to true theories in relevant domains of study (or perhaps a single overarching theory). The results of experimentation are not up to us, but depend on a 'force outside of ourselves', that is, the reality. So theory change is, in the long run, guided by the reality itself.

By analogy, in PP, a model of the environment is acquired through prediction error minimization. This process is guided by the external, mind-independent causal structure. Because there is always some prediction error, the brain fails to achieve a full 'convergence', where it would stabilize on a set of priors which completely prevent it from being surprised by incoming data. However, a suggestion can be made that a partial convergence is achieved in the 'abstract' levels of the generative model which track more time-invariant regularities in input.

Consider the process by which human children develop a mature understanding of the physical object, which then remains stable through the rest of life. If PP is to be believed, this is achieved not by a triggering of an innate representation, but rather through prediction-error-minimization-based learning process. On this story, at a relatively high level of the generative model, a prior is empirically acquired that constrains lower level(s) to parse the sensory stream

by tracking entities that meet the 'folk' criteria of objecthood. Because this prior is induced from the sensory signal, and because it resists further revision as new data come in, learning it counts as the generative model converging on a view of reality as composed of ordinary objects. Furthermore, to learn this prior, the brain relies on converging sensory evidence coming from multiple sense modalities (multisensory correlations), acting as distinct information channels. This is analogous to distinct experimental or detection methods in science corroborating each other.[9] Lastly, the convergence goes beyond processes that take place within a single brain. Arguably, humans share a common, largely cross-cultural understanding of a physical object. On the present view, this means that distinct brains converge on a single way of structuring the environment. This is presumably because acquiring the 'object prior' depends on sensory input produced by a common causal structure (simply, the world), regardless of the differing (for example, cultural) starting points or the details of individual developmental trajectories.

6   Conclusions and some open issues

Let me close by adding some caveats to the foregoing discussion and by sketching out some further issues.

Most importantly, the argument laid out here obviously relies on PP. It also treats PP as committed to internal representation and the inferential view of perception. But as things stand, PP's empirical adequacy, conceptual consistency, explanatory strength, and the

---

[9] Perhaps this point could also be made without recourse to an empiricist view of the origins of the understanding of objects. Natural selection could itself be construed as a sort of Bayesian model selection mechanism (Badcock *et al*. [2019]), which learns the object prior (and 'encodes' it implicitly in the organism's morphology or innate learning biases) over evolutionary timescale. Thus, although PP is naturally interpreted along empiricist lines, it needn't be committed to wholesale rejection of innate representations.

tractability of the computations it postulates are still under question (cf. Litwin and Miłkowski [2020[; Kwisthout and van Rooij [2019]). And even assuming that further development will resolve those issues in PP's favor, it is still not settled whether a representational and inferential interpretation is the right way of understanding the theory (cf. Bruineberg *et al.* [2016]; Orlandi [2018]).

Two considerations mitigate this worry. First, a substantial part of my argument could be grounded in other approaches and theories. The core Bayesian view of perception—of which PP is only a particular incarnation—preserves the idea of object perception as resulting from causal inference. The notion that humans engage in 'epistemic' actions can also be established independently of PP (cf. Cook *et al.* [2011]). To take another example, Green ([2019]) has recently defended the view that object perception picks out stable regularities which are sustained by worldly causes. Green's starting point is the notion that perceptual systems need to encode information efficiently, that is, by compressing it. The idea of efficient coding as central to perception is even more general than Bayesianism itself. Hence, one does not have to buy into PP to accept some of the claims which are of central importance for the present proposal. However, the attraction of PP lies in the fact that it unifies those otherwise scattered ideas under a single view of the brain as minimizing the prediction error. For now, I leave it as an open question just how much of the present case can be salvaged without recourse to PP.

Second, the lesson of the present paper should be taken as a conditional claim. If a certain reading of PP is true, then it constitutes an on-track explanation of object perception. This still seems like a valuable result. It contrasts with the trend of using scientific accounts of the genealogy of intentional states solely for debunking purposes, showing that cognitive details matter in discussions about debunking (Schaffer [2019]).

A further issue worth pointing out is the possibility of extending the present argument by applying it outside the domain of ordinary objects. PP has been employed to explain the origins

of other posits of common sense. For example, Hohwy and Michael ([2017]) argue that the brain infers the self as a hidden, endogenous cause of sensory input. The 'embodied' self explains body-related correlations unfolding over short time-spans, while the 'narrative' self is inferred to explain long-term patterns. Hohwy and Michael take these considerations to favor a realist view of the self. However, Letheby and Gerrans ([2017]), using roughly the same account, arrive at anti-realist conclusions with respect to the self. The present paper suggests that insofar as PP explains the origins of our 'self-models', the realist view is to be favored.

This does not mean that simply by relying on PP, we are always bound to arrive at realist conclusions. Consider the narrative self again. It may be argued that the inferential processes giving rise to a representation of the narrative self are significantly disrupted or skewed by factors that go beyond those invoked by PP (see Williams [2019]). These other factors can be extra-rational (for example, social or motivational) in nature. This leaves room for a fair amount of discrepancy between the selves causing the sensory signals and the selves that people take themselves to be (for example, due to social factors, our brains could dogmatically, often against evidence to the contrary, hold on to a prior belief that we are rational, caring, and moral). More speculatively, perhaps the very propensity to infer a single narrative self as inhabiting a single body is generated by pressures related to efficient social coordination rather than ones related to accurately explaining the sensory input.

Another way of extending the present proposal is by focusing on the idea of rephrasing well-known arguments from philosophy of science so that they apply to perception and its relation to the world. In this paper I focused on arguments that favor realism, but perhaps the same strategy could be employed for revisionary purposes. For illustration, it has been argued that the famous argument from underdetermination of theory by empirical evidence can be restated by treating theory-building in science as solving an inverse problem (Belot [2015]). Roughly, scientists take observable outcomes (experimental data) and construct a theory of

processes (natural laws or causal mechanisms) that generated those outcomes. But, the argument goes, since there isn't always a single unique solution to an inverse problem, multiple empirically equivalent such theories could be possible. By analogy, consider the possibility that there is more than one way the brain could parse the world to effectively minimize the long-term prediction error (note, however, that considerations in Section 3.3 shed doubt on this). Perhaps the prediction error function has multiple local minima and humans are stuck in one of them due to their contingent biological make up, rather than due to convergence to the truth. Some of those alternative predictively successful models might employ strange ontologies that do without ordinary objects. Those models would be equivalent to 'unconceived alternatives'— plausible alternatives to our existing scientific theories which, due to the contingent historical trajectory of science, have never been considered (Stanford [2006]). In any case, while the present discussion aimed to un-debunk tables and chairs, it may open new paths for the debunkers as well.

Paweł Gładziejewski

Nicolaus Copernicus University

Department of Cognitive Science

Toruń, Poland

pawel.gla@umk.pl

References

Adams, W. J., Graf, E. W. and Ernst, M. O. [2004]: 'Experience Can Change the "Light-From-Above" prior', *Nature Neuroscience*, **7**, pp. 1057–1058.

Akins, K. [1996]: 'Of Sensory Systems and the "Aboutness" of Mental States', *The Journal of Philosophy*, **93**, pp. 337–372.

Badcock, P. B., Friston, K. J. and Ramstead, M. J. D. [2019]: 'The Hierarchically Mechanistic Mind: A Free-Energy Formulation of the Human Psyche', *Physics of Life Reviews*, **31**, pp. 104–121.

Belot, G. [2015]: 'Down to Earth Underdetermination' *Philosophy and Phenomenological Research*, **XCI**, pp. 456–464.

Brössel, P. [2017]: 'Rational Relations Between Perception and Belief: The Case of Color', *Review of Philosophy and Psychology*, **8**, pp. 721–741.

Bruineberg, J., Kiverstein, J. and Rietveld, E. [2016]: 'The Anticipating Brain is not a Scientist: The Free-Energy Principle from an Ecological-Enactive Perspective', *Synthese*, **195**, pp. 2417–2444.

Cook, D., Goodman, N. D. and Schulz, L. E. [2011]: 'Where Science Starts: Spontaneous Experiments in Preschoolers' Exploratory Play', *Cognition*, **120**, pp. 341–349.

Clark, A. [2013]: 'Expecting the World: Perception, Prediction and the Origins of Human Knowledge', *Journal of Philosophy,* **CX**, pp. 469–496.

Clark, A. [2016]: *Surfing Uncertainty. Prediction, Action, and the Embodied Mind*, Oxford: Oxford University Press.

Clark, A. [2018]: 'A Nice Surprise?', *Phenomenology and the Cognitive Sciences*, **17**, pp. 521–534.

Dennett, D. [1991]: 'Real Patterns', *Journal of Philosophy*, **88**, pp. 27–51.

Elder, J. H. and Goldberg, R. M. [2002]: 'Ecological Statistics of Gestalt Laws for the Perceptual Organization of Contours', *Journal of Vision*, **2**, pp. 324–353.

Eliasmith, C. [2005]: 'A New Perspective on Representational Problems', *Journal of Cognitive Science*, **6**, pp. 97–123.

Ernst, M. O. and Banks, M. D. [2002]: 'Humans Integrate Visual and Haptic Information in a Statistically Optimal Fashion', *Nature*, **415**, pp. 429–33.

Feldman, J. [2016]: 'The Simplicity Principle in Perception and Cognition', *Wiley Interdisciplinary Reviews – Cognitive Science*, **7**, pp. 330–340.

Forster, M. and Sober, E. [1994]: 'How to Tell When Simpler, More Unified, or Less Ad Hoc Theories Will Provide More Accurate Predictions', *British Journal for the Philosophy of Science*, **45**, pp. 1–36.

Freeman, W. T. [1996]: 'Exploiting the Generic Viewpoint Assumption', *International Journal of Computer Vision*, **20**, pp. 243–261.

Friston, K. J. [2005]: 'A Theory of Cortical Responses', *Philosophical Transactions of the Royal Society B*, **360**, pp. 815–836.

Friston, K. J., Adams, R. A., Perrinet, A. and Brakspear, M. [2012]: 'Perceptions as Hypotheses: Saccades as Experiments', *Frontiers in Psychology*, **3**, 151.

Geisler, W. S. [2008]: 'Visual Perception and the Statistical Properties of Natural Scenes', *Annual Review of Psychology*, **59**, pp. 167–192.

Ghijsen, H. [2018]: Predictive Processing and Foundationalism About Perception, *Synthese*, doi: https://doi.org/10.1007/s11229-018-1715-x.

Gigerenzer, G. and Brighton, H. [2009]: 'Homo Heuristicus: Why Biased Minds Make Better Inferences', *Topics in Cognitive Science*, **1**, pp. 107–143.

Głdziejewski, P. [2016]: 'Predictive Coding and Representationalism', *Synthese*, **193**, pp. 559–582.

Głdziejewski, P. [2017]: 'Evidence of the Senses: A Predictive Processing-Based Take on the Sellarsian Dilemma', in T. Metzinger and W. Wiese (*eds*), *Philosophy and Predictive Processing*, MIND Group, available online at: <https://predictive-mind.net/papers/the-evidence-of-the-senses>.

Głdziejewski, P. and Miłkowski, M. [2017]: 'Structural Representations: Causally Relevant and Different from Detectors', *Biology Philosophy*, **32**, pp. 337–355.

Goldman, A. [1987]: 'Cognitive Science and Metaphysics', *Journal of Philosophy*, **84**, pp. 537–544.

Goldman, A. [1993]: *Philosophical Applications of Cognitive Science*, Boulder, CO: Westview Press.

Green, E. J. [2019]: 'A Theory of Perceptual Objects', *Philosophy and Phenomenological Research*, **99**, pp. 663–693.

Gregory, R. L. [1980]: 'Perceptions as Hypotheses', *Philosophical Transactions of the Royal Society B: Biological Sciences*, **290**, pp. 181–97.

Hacking, I. [1983]: *Representing and Intervening*, Cambridge: Cambridge University Press.

Helmholtz, H. von, [1867]: *Handbuch der Physiologishen Optik*, Leipzig, Leopold Voss.

Hoffman, D. D., Singh, M. and Prakash, C. [2015]: 'The Interface Theory of Perception', *Psychonomic Bulletin & Review*, **22**, pp. 1480–1506.

Hohwy, J. [2013]: *The Predictive Mind*, Oxford: Oxford University Press.

Hohwy, J. and Michael, J. [2017]: 'Why Should Any Body Have a Self?', in F. de Vignemont and A. Alsmith (*eds*), *The Subject's Matter: Self-consciousness and the Body*, Cambridge (MA): The MIT Press, pp. 363–392.

Kemp, C., Perfors, A. and Tenenbaum, J. B. [2007]: 'Learning Overhypotheses with Hierarchical Bayesian Models', *Developmental Science*, **10**, 307–321.

Kiefer, A. [2017]: 'Literal Perceptual Inference', in T. Metzinger and W. Wiese (*eds*), *Philosophy and Predictive Processing*, MIND Group, available online at: <https://predictive-mind.net/papers/literal-perceptual-inference>.

Kiefer, A. and Hohwy, J. [2017]: 'Content and Misrepresentation in Hierarchical Generative Models', *Synthese*, **195**, pp. 2387–2415.

Korman, D. Z. [2015]: *Objects: Nothing Out of the Ordinary*, Oxford: Oxford University Press.

Korman, D. Z. [2019]: 'Debunking Arguments', *Philosophy Compass*, **14**, e12638.

Körding, K. P., Beierholm, U., Ma, W. J., Quartz, S., Tenenbaum, J. B. and Shams, L. [2007]: 'Causal Inference in Multisensory Perception', *PLOS One*, **9**, e943.

Kwisthout, J. and van Rooij, I. [2019]: 'Computational Resource Demands of a Predictive Bayesian Brain', *Computational Brain Behavior*, **3**, pp. 174–188.

Ladyman, J. and Ross, D. [2007]: *Everything Must Go: Metaphysics Naturalized*, Oxford: Oxford University Press.

Letheby, C. and Gerrans, P. [2017]: 'Self Unbound: Ego Dissolution in Psychedelic Experience', *Neuroscience of Consciousness*, **2017**, 1–11.

Lee, J. [2018]. 'Structural Representation and the Two Problems of Content', *Mind & Language*, **34**, pp. 606–626.

Litwin, P. and Miłkowski, M. [2020]: 'Unification by Fiat: Arrested Development of Predictive Processing', *Cognitive Science*, doi: https://doi.org/10.1111/cogs.12867

McKay, R. T. and Dennett, D. [2009]: 'The Evolution of Misbelief', *Behavioral and Brain Sciences*, **32**, pp. 493–510.

Munton, J. [2018]: 'The Eye's Mind: Perceptual Process and Epistemic Norms', *Philosophical Perspectives*, **31**, pp. 317–347.

Orlandi, N. [2018]: 'Predictive Perceptual Systems', *Synthese*, **195**, pp. 2367–2385.

Osborne, R. C. [2016]: 'Debunking Rationalist Defenses of Common-Sense Ontology: An Empirical Approach', *Review of Philosophy and Psychology*, **7**, pp. 197–221.

Parise, C. V., Spence, C. and Ernst, M. O. [2012]: 'When Correlation Implies Causation in Multisensory Integration', *Current Biology*, **22**, pp. 46–49.

Peirce, C. S. [1878/2011]: 'How to Make Our Ideas Clear', in R. B. Talisse and S. F. Aikin (*eds*), *The Pragmatism Reader: From Peirce Through the Present*, Princeton (NJ): Princeton University Press, pp. 50–65

Perfors, A., Tenenbaum, J. B., Griffiths, T. L. and Xu, F. [2011]: 'A Tutorial Introduction to Bayesian Models of Cognitive Development', *Cognition*, **120**, pp. 302–321.

Putnam, H. [1975]: *Mathematics, Matter and Method*, Cambridge: Cambridge University Press.

Schaffer, J. [2019]: 'Cognitive Science and Metaphysics: Partners in Debunking', in A. Goldman and B. McLaughlin (*eds*), *Metaphysics and Cognitive Science*, Oxford: Oxford University Press.

Shams, L. and Beierholm, U. [2010]: 'Causal Inference in Perception', *Trends in Cognitive Sciences*, **14**, pp. 425–432.

Siegel, S. [2017]: *The Rationality of Perception*, Oxford: Oxford University Press.

Sigman, M., Cecchi, G. A., Gilbert, C. D. and Magnasco, M. O. [2001]: 'On a Common Circle: Natural Scenes and Gestalt Rules', *Proceedings of the National Academy of Sciences*, **98**, pp. 1935–1940.

Sotiropoulos, G., Seitz, A. R. and Seriès, P. [2011]: 'Changing Expectations About Speed Alters Perceived Motion Direction', *Current Biology*, **21**, R883–R884.

Spelke, E. S. [1990]: 'Principles of Object Perception', *Cognitive Science*, **14**, pp. 29–56.

Stanford, K. [2006]: *Exceeding Our Grasp: Science, History, and the Problem of Unconceived Alternatives*, Oxford: Oxford University Press.

Treisman, A. and Schmidt, H. [1982]: 'Illusory Conjunctions in the Perception of Objects', *Cognitive Psychology*, **14**, pp. 107–41.

Wiese, W. [2018]: *Experienced Wholeness: Integrating Insights from Gestalt Theory, Cognitive Neuroscience, and Predictive Processing*, Cambridge, MA: The MIT Press.

Williams, D. [2017]: 'Predictive Processing and the Representation Wars', *Minds and Machines*, **28**, pp. 141–172.

Williams, D. [2019]: 'Epistemic Irrationality in the Bayesian Brain', *British Journal for the Philosophy of Science*, DOI: https://doi.org/10.1093/bjps/axz

Witkin, A.P. and Tenenbaum, J.M. [1983]. 'On the Role of Structure in Vision, in J. Beck, B. Hope and A. Rosenfeld (*eds*), *Human and Machine Vision*, New York: Academic Press, pp. 481–543.