

Bayesian perception is ecological perception*

Nico Orlandi

There is a certain excitement in vision science concerning the idea of applying the tools of Bayesian decision theory to explain our perceptual capacities. Bayesian models – or ‘predictive coding models’ – are thought to be needed to explain how the inverse problem of perception is solved, and to rescue a certain constructivist and Kantian way of understanding the perceptual process (Clark 2012, Gladziejewski 2015, Hohwy 2013, Rescorla 2013, Rescorla forthcoming.)

Anticlimactically, I argue both that Bayesian outlooks do not constitute good solutions to the inverse problem, and that they are not constructivist in nature. In explaining how visual systems derive a single percept from underdetermined stimulation, orthodox versions of predictive coding accounts encounter a problem. The problem shows that such accounts need to be grounded in Natural Scene Statistics (NSS), an approach that takes seriously the Gibsonian insight that studying perception involves studying the statistical regularities of the environment in which we are situated.

Additionally, I argue that predictive coding frameworks postulate structures that hardly rescue a constructivist way of understanding perception. Safe for percepts, the posits of Bayesian theory are not representational in nature. Bayesian perceptual inferences are not genuine inferences. They are biased processes that operate over non-representational states. In sum, Bayesian perception is ecological perception.

In section 1, I present what Bayesianism about perception amounts to. In section 2, I argue that Bayesian accounts, as standardly described, do not offer a satisfactory explanation of how the inverse problem is solved. Section 3 introduces Natural Scene Statistics and argues that predictive coding approaches should be combined with it. Section 4 argues that the posits of Bayesian theory are in line with an ecological way of understanding the

*This article develops an argument that I first attempted in Orlandi 2014. I thank the audience of the 2015 Barnard-Rutgers-Columbia Mind workshop for useful comments.

perceptual process. Section 5 closes with a brief historical aside that aims to make the main claims of this article more palatable.

1 Bayesian perception

Bayesian models are often introduced by noticing that perceptual systems operate in conditions of uncertainty and Bayesian decision theory is a mathematical framework that models precisely decision-making under uncertainty (Rescorla 2013). In vision, one important source of uncertainty is the presumed ambiguity of sensory input. The stimulus that comes to our sensory receptors in the form of a pattern of light is said to underdetermine its causes, in the sense of being compatible with a large number of different distal objects and scenes.

In one of David Marr’s examples, a discontinuity in light intensity at the retina could be caused by many distal elements (Marr and Hildreth 1980). It could be caused by edges, changes in illumination, changes in color, cracks. Indeed, “there is no end to the possible causes” (Hohwy 2013, p. 15).

We do not, however, see the world in a constantly shifting way. We typically see one thing in a stable way. We need a story of how the visual system – usually assumed to comprise retina, optic nerve and visual cortex – derives a single percept from ambiguous retinal projections. This has come to be known as the ‘inverse problem’, or the ‘underdetermination problem’ (Howhy et al. 2008, Palmer 1999, Rock 1983).

Bayesian accounts hold that the inverse problem is solved by bringing to bear some implicit knowledge concerning both what is more likely present in the environment (prior knowledge) and what kind of distal causes gives rise to a given retinal projection (knowledge of likelihoods). Adams and his collaborators, for example, say:

To interpret complex and ambiguous input, the human visual system uses prior knowledge or assumptions about the world.
(Adams et al. 2004, p. 1057)

Such prior knowledge is sometimes described as being a body of *beliefs* or *expectations*. Hohwy, for example says:

In our complex world, there is not a one-one relation between causes and effects, different causes can cause the same kind of

effect, and the same cause can cause different kinds of effect. (...) If the only constraint on the brain's causal inference is the immediate sensory input, then, from the point of view of the brain, any causal inference is as good as any other. (...) It seems obvious that causal inference (...) draws on a vast repertoire of prior belief. (Hohwy 2013, p. 13)

Since visual processing is a sub-personal affair, the repertoire of knowledge introduced by Bayesian frameworks – even when described as comprising beliefs – is usually not available to consciousness. In this sense, it is implicit. Mamassian and his collaborators clarify this point:

We emphasize the role played by prior knowledge about the environment in the interpretation of images, and describe how this prior knowledge is represented as prior distributions in BDT. For the sake of terminological variety, we will occasionally refer to prior knowledge as “prior beliefs” or “prior constraints”, but it is important to note that this prior knowledge is not something the observer need be aware of. (Mamassian et al.2002, p.13)

In employing implicit knowledge the perceptual process is said to resemble a process of unconscious reasoning – that is, a process akin to an inference. Accordingly, Gladzeijewski says:

PCT rests on the general idea that in order to successfully control action, the cognitive system (the brain) has to be able to infer “from the inside” the most likely worldly causes of incoming sensory signals. (...) Such a task is unavoidably burdened with uncertainty, as no one-to-one mapping from worldly causes to sensory signals exists. (...) According to PCT, the brain deals with this uncertainty by implementing or realizing (approximate) Bayesian reasoning. (Gladzeijewski 2015, p. 3.)

To better detail this framework, we can consider the example of shape perception (Brainard 2009, Hosoya et al.2005). The configuration in figure 1 is typically perceived as consisting of convex half-spheres illuminated from above. The configuration, however, is also compatible with the presence of concave holes illuminated from below (Ramachandran 1988).

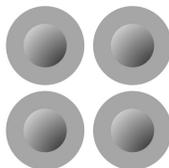


Figure 1: Shaded dots typically seen as convex. Based on Ramachandran 1988.

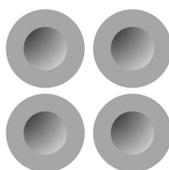


Figure 2: Shaded dots typically seen as concave. Based on Ramachandran 1988.

You may be able to see the spheres as holes on your own. Figure 2 helps by showing that rotating the configuration makes the central elements appear concave.

Given that the stimulus produced by figure 1 is compatible with different percepts, the question is why we typically see it in just one way. In a Bayesian framework this question is answered in roughly the following manner. The perceptual system starts by evaluating a set of hypotheses – the so-called ‘hypotheses space’, which comprises ‘guesses’ as to what is in the world. In this case, we can think of the system as considering two hypotheses: the hypothesis that convex half-spheres illuminated from above are present (let’s call this hypothesis S) and the hypothesis that concave holes illuminated from below are present (let’s call this hypothesis C).

One hypothesis from this space is assigned higher initial probability. This is because of the implicit knowledge, or assumptions that bayesian perceptual systems are thought to store. Human vision assumes that light comes from a single source and from *above*. It tends to presume that things are illuminated by something overhead, like the sun (Mamassian et al. 2002, Ramachandran 1988, Rescorla 2013, Stone 2011, p.179). Moreover – although this is still controversial – perceptual systems presented with specific shapes are said to prefer convexity to concavity.¹ This makes it so that hypothesis S has higher

¹For discussion of this prior, see Stone 2011, p. 179.

initial probability than C (so $p(S) > p(C)$).

Hypothesis S is then checked against the evidence, that is, against the incoming sensory input. It is rather unclear how this checking takes place. A plausible way of understanding Bayesian models is that the hypothesis generates a mock stimulus – that is, a simulation of what the sensory input would have to be like if the hypothesis were true. This mock stimulus is matched against the actual sensory input and an error signal is generated if there is mismatch. This overall process produces a measure of the ‘likelihood’ of the hypothesis, that is of how likely the hypothesis is given the sensory evidence.

In order to calculate the likelihood function, perceptual systems use again some prior ‘knowledge’. This is knowledge of the kind of sensory stimulation that certain distal causes produce. The visual system must know what kind of retinal configuration is typically caused by convex half-spheres illuminated from above. It uses this knowledge to test the initial hypothesis (Beierholm et al. 2009).

In figure 1, the sensory stimulation is compatible with hypothesis S , so the error signal is minimal or non-existent, and the likelihood is correspondingly high. The likelihood measure together with the prior probability is then used to calculate the posterior probability of the hypothesis – which is proportional to the product of prior and likelihood – and the percept is selected as a function of this probability. S is selected because it is probable, and it does not produce a significant error signal. Figure 3 is a schematic illustration of the whole process.

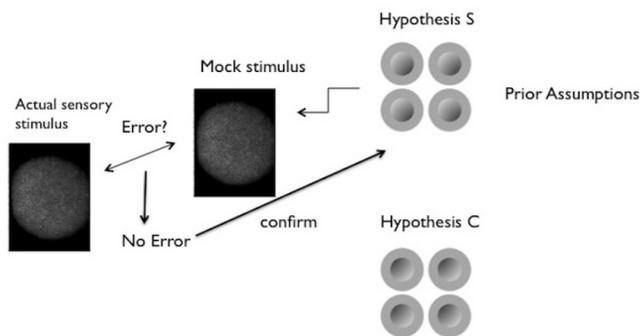


Figure 3: Simplified illustration of hypothesis-testing in vision. The sensory stimulus is a pattern of light projected on the retina.

If there is error, or ‘surprisal’, the brain may respond in a number of ways. It may reject the initial hypothesis as the percept. It may adjust it ‘from the inside’ to better match what is in the world. Or it may produce *action* that changes the world to adjust to the hypothesis. In each case, the basic insight is that error is discomforting, and the brain tries to reduce it.

In some Bayesian models, the final selection of the percept is made in accord with expected utility maximization which involves calculating the costs and benefits associated with accepting the hypothesis with the highest posterior. In others, like the one I just described, it is made simply as a function of the posterior probability (Maloney and Mamassian 2009, Howe et al. 2008, p. 2; Mamassian et al. 2002, p. 20–21, Mamassian and Landy 1998).²

The implicit assumptions used to select the hypotheses from the hypothesis space are called ‘priors’ and ‘hyperpriors’. The assumption that things are illuminated from above, being pretty general, is typically regarded as a hyperprior (Howhy et al. 2008, p. 5, Mamassian et al. 2002). Preference for convexity when it comes to specific objects, by contrast, is regarded as a prior. There is disagreement concerning whether priors and hyperpriors are innately specified or learnt from experience (Beierholm et al. 2009, p. 7, Brainard 2009, p. 25, Clark 2012 footnote xxxvi, Hohwy et al. 2008, p. 3–4, Mamassian et al. 2002, p. 13).

As it is evident from this simplified example, implicit knowledge is central in Bayesian accounts of perception. The sensory evidence in figure 1 is compatible with both hypothesis *S* and hypothesis *C*. Accordingly, the likelihood of hypothesis *S* is the same as the likelihood of hypothesis *C* ($p(l/S) = p(l/C)$). Because the priors assign higher probability to *S*, however, this hypothesis ends up having higher posterior, and the shapes in figure 1 are seen as convex.

²The Bayesian explanation that I just outlined is simplified in several respects. For one thing, I described predictive coding models as composed of only three levels – the high-level perceptual hypothesis, the mock sensation and the actual sensation. Bayesian inferences, however, are highly hierarchical with several intermediate steps, a point on which we will return. I also described the visual apparatus as operating over discrete hypotheses, rather than over probability density functions (Clark 2012, p. 22). Further, I worked under the idealization that vision is an isolated process, while there are certainly influences from other modalities. I also left aside the issue of how faithful these models are to the notions of Bayesian decision theory. Interested readers can refer to the discussion in Mamassian et al. 2002, p. 29. I do not think that these simplifications perniciously affect the main argument of this paper.

So the priors and hyperpriors have a significant role in this context. In other contexts, the likelihood might ‘beat’ the prior and a hypothesis that has relatively low initial probability may be selected. This is how we are presumably able to perceive surprising things. Provided that the prior probability is high enough, however, perceivers see just what they *expect* to see.

Now, despite the apparent plausibility of this view, I think that Bayesian accounts provide only a *prima facie* explanation. The machinery that they introduce is not crucial to explain why we typically see just one thing in response to retinal projections. The next section explains why this is the case.

2 A puzzle for Bayesian inferences

Standard presentations of Bayesian models focus overwhelmingly on the role that priors and hyperpriors play in reducing uncertainty. The underlying thought seems to be that priors and hyperpriors guarantee that we see the world in a unique way because they constitute prior knowledge that selects one percept over others that are also compatible with the stimulus. Indeed, some proponents of predictive coding models seem to want to reconceptualize the importance of the incoming input. Perception is driven ‘from the top’, by prior knowledge of the world. Incoming sensory signals are responsible mostly for detecting error. Gladzejewski, for example writes:

PCT presents us with a view of perception as a Kantian in spirit, “spontaneous” interpretive activity, and not a process of passively building up percepts from inputs. In fact, on a popular formulation of PCT, the bottom-up signal that is propagated up the processing hierarchy does not encode environmental stimuli, but only signifies the size of the discrepancy between the predicted and actual input. On such a view, although the world itself surely affects perception – after all, the size of the bottom-up error signal is partly dependent on sensory stimulation – its influence is merely corrective and consists in indirectly “motivating” the system, if needed, to revise its perceptual hypotheses. (Gladzejewski 2015, p. 16)

Hohwy similarly writes:

Perception is more actively engaged in making sense of the world than is commonly thought. And yet it is characterized by curious passivity. Our perceptual relation to the world is robustly guided by the offerings of the sensory inputs. And yet the relation is indirect and marked by a somewhat disconcerting fragility. The sensory input to the brain does not shape perception directly: sensory input is better and more perplexingly characterized as feedback to the queries issued by the brain. Our expectations drive what we perceive and how we integrate the perceived aspects of the world, but the world puts limits on what our expectations can get away with. (Hohwy 2013, p. 2)

The error detection function of sensory states is what makes Bayesian accounts into predictive coding accounts. Predictive coding is originally a data-compression strategy based on the idea that transmitting information can be done more effectively if signals encode only what is unexpected. An image, for example, can be transmitted by a code that provides information about boundaries, rather than about regions that are highly predictable given the surrounding context.³ Like in predictive coding, sensory states provide feedback if there is error in the current hypothesis.

Upon reflection, however, the focus on expectations and the stress on the error-detection function of sensory inputs is misguided. What must play a crucial role in explaining why we usually see just one thing is precisely the incoming sensory data that is presumed to be ambiguous.

To see why this is the case, consider a traditionally hard question for Bayesian models in general (not just for Bayesian models of perception). The question is how the initial hypothesis space is restricted.

In typical presentations of the predictive coding approach, the brain assigns prior probabilities to a set of hypotheses (Mamassian et. al 2002, p. 14; Rescorla 2013, p. 5). In the case of shape perception in figure 1, vision starts out by evaluating two hypotheses: that convex half-spheres are present and that concave holes are present. What is hardly discussed, however, is that, given genuinely underdetermined stimuli, the hypothesis space is infinite and we lack a story of how it is restricted (Rescorla 2013 p. 4). Why does the visual system consider hypotheses about half-spheres and holes, rather than about something else? How does it know that what is present is either a half-sphere or a hole?

³See, Clark 2012 p. 5, Shi and Sun 1999.

These questions are sometimes brushed aside in order to get the explanation going, but they cannot be ignored in the present dialectical situation. If we are seeking to explain how we derive a single percept from underdetermined stimuli, then we cannot leave aside the question of how the hypothesis space is limited. This would amount to trading the original mystery with a new, similar mystery.

The question of how to restrict the hypothesis space is a question of the same kind as the inverse problem question. Both questions concern how to reduce uncertainty. Both questions are about how to get to one leading hypothesis, or to a few leading hypotheses given underdetermined data. Both questions need to be answered in order to get a satisfactory explanation.

This is not always true. Sometimes it is possible to explain X by citing Y without in turn explaining Y (Rescorla 2015). To quote Rescorla:

For example, a physicist can at least partially explain the acceleration of some planet by citing the planet's mass and the net force acting on the planet, even if she does not explain why that net force arises.

Although this is in general a good point, this analogy does not quite work. Explaining how uncertainty is reduced by appeal to the notion of a hypothesis space whose uncertainty we do not know how to reduce, would seem to just fail to explain the initial datum. A more fitting analogy would be if we tried to explain acceleration by appeal to a notion that is itself like acceleration, but perhaps called with a different name.

We need some sort of story of how the hypothesis space is restricted, and I do not think that Bayesian models provide a story. That is, they do not provide a story *unless* they recognize that the incoming signal has a much bigger role than priors and hyperpriors in achieving a single percept. The idea would be that the configuration in figure 1 projects on the retina a pattern of stimulation that, in our world, is often caused by either half-spheres illuminated from above or holes illuminated from below. This is why these two hypotheses enter the restricted hypothesis space.

This proposal gains credibility from the fact that priors and hyperpriors do not do the required work in this context. Hyperpriors are too general to be helpful. Hyperpriors are assumptions concerning light coming from above or a single object occupying a single spatio-temporal location. As such, they are not able to restrict the initial set to hypotheses about particular objects or particular configurations.

Additionally, both priors and hyperpriors concern a variety of things. To play the appropriate role they have to be activated in the relevant situations. One example of a prior, as we saw, is the assumption that objects tend to be convex. Another example, that I mention in the next section, is the assumption that edge elements that are oriented roughly in the same way belong to the same physical edge. Priors concern many different elements, presumably as many as the variety of objects and properties that we can perceive. They concern color, faces and horizontal structure in the world (Geisler et al. 2001, p. 713, Geisler 2008, p. 178).

A prior about horizontal structure would not be of much use in disambiguating figure 1. In a sense, the brain has to be ‘told’ that it is the prior concerning convexity that is relevant, rather than some other prior. But it seems that the only candidate to tell the brain any such thing is the driving signal that reflects what is in the world. The signal must activate the kinds of priors and hyperpriors that should be employed in a given instance.

Some hyperpriors may of course be active in every context. For example, the hyperprior concerning light coming from above may be active *whenever we see*. But hyperpriors of this kind are again too general to sufficiently constrain the perceptual hypotheses that are relevant. Given the variety of objects and properties that we can see, a hyperprior concerning illumination would not be able to sort through what specific kinds of objects we are likely to perceive in a particular instance.

Would appeal to the likelihood function help? Perhaps only the hypotheses that are likely given the sensory input enter the restricted hypothesis space. Although this is a plausible suggestion, it is not in step with standard presentations of Bayesian models. The likelihood function typically enters the scene when the currently winning hypothesis is being tested. It is used to produce mock sensory stimulation, not to restrict the initial set.

What is distinctive of Bayesian accounts is that priors and hyperpriors *guide* the perceptual process. This is what distinguishes such accounts from ‘bottom-up’ inferential views that I briefly describe in section 4. Accordingly, proponents of predictive coding views tend to think that, even factoring-in likelihood, the information coming from the world is ambiguous. Hohwy, for example, says:

We could simplify the problem of perception by constraining ourselves to just considering hypotheses with a high likelihood. There will still be a very large number of hypotheses with a high

likelihood simply because, as we discussed before, very many things could in principle cause the effects in question. (Hohwy 2013, p. 16)

Given highly underdetermined input, very many hypotheses enjoy high likelihood. Indeed if the likelihood played a larger role, then it would be unclear why we would need priors and hyperpriors after all. Why not just select the hypothesis that is the most likely? This, in effect, would amount to conceding that the bottom-up signal is constraining in various ways.

We are then back to our initial problem. If the hypothesis space is very large, we need a story of how to restrict it. It seems that the only plausible candidate to do the restricting job is the information coming from the bottom, or from the world. This information makes the hypothesis space manageable, and it activates the priors that are relevant in a given context. Predictive coding accounts seem bound to admit that priors and hyperpriors do not in fact play a central role in reducing uncertainty. The natural alternative is that the driving signal does.

We now have all the elements to question Bayesianism as a needed framework to solve the inverse problem. If we are open to appealing to the driving signal to diminish uncertainty from infinity to a very limited set of hypotheses, why not also think that the driving signal reduces uncertainty altogether? Why think that priors and hyperpriors are needed at all? The driving signal does not just restrict the hypothesis space. It also tells the visual system that the stimulus is more likely caused by something convex. It reduces uncertainty not just to a hypothesis space, but to *one* leading hypothesis. Its statistical properties make it so that the configuration in figure 1 is first seen as half-spheres. If we rotate the image, the statistical properties change, making it so that a different hypothesis is formed. In this picture, vision forms hypotheses about half-spheres because it is exposed to them (or to something that looks like them) not because of an internal inferential process.⁴

Accepting this point amounts to recognizing, in Gibsonian spirit, that the stimulus for perception is richer than it is supposed in standard presentations of the inverse problem. It is a stimulus that is underdetermined only in principle, or in highly artificial or ‘invalid’ experimental conditions. It is not a stimulus that is underdetermined in our specific environment. In such

⁴Some proponents of Bayesian models accept this point and are sympathetic to the view that predictive coding models have to be grounded in Natural Scene Statistics. Clark (personal communication). See also 2012, p. 28–29 and p. 46.

environment, the stimulus is highly constraining, limiting the number of cerebral configurations (or hypotheses) that are appropriate.

To summarize: I started this section by raising a problem for Bayesian ways of accounting for how we derive a single percept from ambiguous stimulation, namely that we have no story of how the hypothesis space is restricted. Bayesians owe us an explanation of why, and how, an initial set is formed given that the information coming from the world is supposedly ambiguous and presumably un-constraining. I then suggested that the most natural way to explain how the hypothesis space is restricted is by appeal to the driving signal. Why are some hypotheses and not others winning in some contexts? The natural answer is that it is because of the context. This suggests that the emphasis on perception as a process guided by expectations, and the reconceptualization of sensory states as primarily error detectors is misguided. Predictive coding models have to be grounded in a study of the environmental conditions in which vision occurs.

From here, I suggested that, if we can appeal to the driving signal to restrict the hypothesis space, we can appeal to it also to explain the selection of the winning hypothesis from the set. The Bayesian machinery of priors, hyperpriors and even likelihoods does not seem to be needed to reduce uncertainty. It may be needed for other purposes (more on this in section 3), but it is not needed for the selection of the initial perceptual guess, which is pretty bad if that is what we introduced the models to do.

We should then re-consider Bayesianism as a solution to the inverse problem. What guides perception are not expectations and prior beliefs, but the driving stimulation. To understand how this happens we should turn to a research program in perceptual psychology called Natural Scene Statistics (NSS).

3 Natural scene statistics

Natural scene statistics is an approach that originates in physics⁵ and that enjoyed some recent fortune due to the improved ability of computers to parse and analyze images of natural scenes. NSS comes in different forms: some of its practitioners are sympathetic to Bayesian accounts of perception (Geisler 2008) while others are critical (Howe et al. 2006, Yang and Purves

⁵Geisler 2008 p. 169.

2004). For present purposes, I leave this disagreement aside and outline the features of NSS that are relevant to the underdetermination problem.

One of the fundamental ideas of NSS is to use statistical tools to study not what goes on inside the head, but rather what goes on outside. NSS is interested both in what is more likely present in our environment, and in the relationship between what is in the world and the stimulus it produces.

Researchers in this paradigm often take a physical object or property and measure it in a large number of images of natural scenes. We can calculate the incidence of a certain luminance in natural environments (Howe et al. 2006, Yang and Purves 2004), or the incidence and characteristics of edge elements in the environment (Geisler et al. 2001).

When we proceed in this way we discover facts about the world that we can later use to understand perception. We may find, for example, that edge elements are very common in our world, and that, if the elements are oriented more or less in the same way, they are more likely to come from the same physical edge. This reflects the fact that natural contours have relatively smooth shapes (Geisler 2008, p. 178).

In addition to calculating what is common in the world, proponents of NSS can also calculate the kind of input that environmental elements project on the retina. This is typically done by measuring the probability of a given retinal stimulation being caused by a particular item. For example, by calculating what type of projection oriented edge elements produce on the retina, we can predict whether any two elements will be seen as belonging to the same physical contour. Since we know that edge elements oriented in the same way typically come from the same edge, we can expect perceivers to see them that way.

NSS discovers the law-like or statistical regularities that priors are supposed to mirror (Rescorla, forthcoming p. 14). By doing so, it makes appeal to the priors themselves superfluous. It is because of the regularities themselves, not because of prior knowledge of them, that we see continuous edges.

In this framework, a retinal projection gives rise to the perception of a single contour not because of an inference, but because a single contour is the most likely cause of the given retinal state. Indeed, it would be surprising if it were otherwise. Since the world we inhabit has certain discoverable regularities that may matter to us as biological systems, it would be surprising if we were not equipped to respond to them at all.

A more sophisticated example may help further clarify this way of approaching vision. The intensity of light reflecting from objects in the spec-

trum that our eyes are able to respond to is typically called ‘luminance’ (Purves and Lotto 2003, chapter 3). Luminance is a physical quantity measurable in natural scenes. An increase in luminance means an increase in the number of photons captured by photoreceptors and an increase in their level of activation.

We may intuitively suppose that the perception of brightness – that is, how bright or light we perceive something to be, is a direct measure of luminance. But this is famously not true. Target patches that are equi-luminant appear different in brightness depending on context. In figure 4, the target dot with light surround appears darker than the equi-luminant target dot with dark surround.

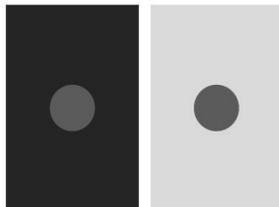


Figure 4: Brightness contrast effect from Purves and Lotto 2003 p. 56.

Luminance and brightness perception are not directly proportional because, in natural environments, luminance is a function not just of the reflectance of surfaces – that is, of the quantity of light reflected by an object – but also a function of both their illumination (the amount of light that hits them) and the transmittance properties of the space between the objects and the observer – for example, the atmosphere (Howe et al. 2006). For simplicity, we can focus here on reflectance and illumination as the two primary variables that determine luminance. The same luminance can be caused by a dark surface illuminated brightly, and by a less illuminated light surface. Vision has to somehow disambiguate this contingency and figure out what combination of reflectance and illumination is responsible for a given luminance quantity at the retina.

In natural scene statistics, solving this problem means looking at the environmental conditions that give rise to perceived brightness. The target circle on the left of figure 4 appears lighter than the target circle on the right despite the fact that the two targets are equiluminant. In explaining why this is the case, we start by analyzing a large database of natural images. In particular, we can superimpose the configuration in figure 4 on a large number

of images to find light patterns in which the luminance of both the surround and the target regions are approximately homogeneous to the elements of the figure (Yang and Purves 2004, p. 8746). This gives us a sense of what typically causes the configuration in question.

Doing this, we discover that, in our world, high-luminance surrounds co-occur more frequently with high-luminance targets than with low-luminance targets. *Vice versa* low-luminance surrounds co-occur more frequently with low-luminance targets. This means that the stimuli in figure 4 are more likely caused by the scene in the *right* image of figure 5. They are more likely caused by targets that are both illuminated differently and that have different reflectance properties.

The luminance of the two center targets in the figure is the same. Given our environment, however, the target that is embedded in a lighter surround should be more luminant. The fact that it is not suggests that it is darker – that is, it suggests that it is illuminated more brightly, and that it reflects less light. Similarly, the target on the left of the figure should be less luminant. The fact that it is not suggests that it is brighter – that is, it suggests that it has more reflectance but it is illuminated by less light. The targets are accordingly seen as differing in brightness.⁶

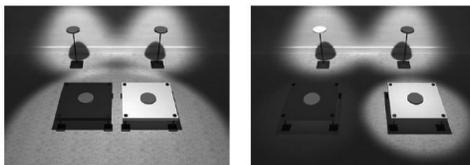


Figure 5: Two possible sources of the brightness contrast effect from Purves and Lotto 2003 p. 56. The right scene is, in our environment, the more common cause of the stimulus.

Again, in accounting for brightness perception we do not appeal to an inference, but to the usual causes of a given configuration. If the retinal stimulus is typically produced by things that are different in brightness, then we are likely to see them that way.

What we are doing is essentially applying statistical tools to study the environmental conditions in which vision occurs. And if we do that, we

⁶Howe et al. 2006, p. 872. This approach is also able to explain a number of brightness effects including the Cornsweet edge effects and Mach bands (Purves and Lotto 2003, ch. 4).

develop a powerful method to predict what observers see. In particular, we discover two things. First, the stimulus for vision is much more constraining than what we first thought. Second, given the statistical properties of the stimulus, it is unsurprising that vision would respond to certain conditions with some hypotheses and not others. So we explain how the inverse problem is solved by appeal to how the world is.

If this is true, then the Bayesian machinery seems superfluous to solve the underdetermination problem. Appeal to priors and hyperpriors is of course not excluded by appeal to stimulus conditions. We could appeal to *both* in a full explanation of how uncertainty is reduced.⁷ But while this is, in general, a point well-taken, it is somewhat strange in the present context. Predictive coding accounts claim that we need to posit implicit assumptions and expectations in order to explain how a single percept is secured (see section 1). If we then show that a unique percept is guaranteed by other means, it would seem to follow that the Bayesian machinery is explanatorily redundant – that is, *in the context of explaining how we derive a single percept* we do not need to introduce it at all. This is not without irony since Bayesian approaches are thought to be particularly well-appt at explaining perception (Hohwy 2013, p. 7).

But now, suppose that one agrees that the driving signal has a big role to play in vision. One can be in favour of grounding predictive coding accounts in Natural Scene Statistics, but think that the Bayesian machinery is still needed. This may be the case for various reasons. One may think that the emphasis on the inverse problem is misguided. Bayesian models describe the kinds of structures that the brain grows when it attunes to environmental contingencies. Such models describe the kinds of assumptions that the brain develops as a result of being situated in our world.

In so far as we can expect the brain to develop tendencies to presume, for example, that light comes from above, or that objects have smooth, continuous edges, we can accept Bayesian frameworks as descriptions of these tendencies. The descriptions, in turn, can prove useful in explaining other aspects of perception: its phenomenology, its constancy, its coherence despite information coming from seemingly separate sensory channels (Hohwy 2013).

The insight that the brain retains a hypothesis through time, provided that the error signal is not too significant, for example, explains diachronic aspects of perception. When we see something, we continue to see it more

⁷Thanks to Michael Rescorla for pressing this point in Rescorla 2015.

or less unaltered, despite changes in the light the object reflects.

In the rest of this paper, I argue that even if one accepts this position, one is still within the bounds of a Gibsonian and ecological way of modelling perception. This is because, save for perceptual high-level hypotheses, the postulates of Bayesian perceptual models are not representational in nature. Despite appearances, Bayesian outlooks do not rescue constructivism and Kantianism in perceptual psychology.

4 Bayesian brains are ecological brains

A long standing dispute in psychological theories of perception is that between inferentialism and ecology (Helmholtz 1867/1925, Rock 1983, Gibson 1966). Theories that belong to the former camp tend to view perception as an indirect relation to the world, and to consist in a process that involves an internal construction of distal objects and scenes. As we saw, this construction is often understood as a type of inference that makes use of internally stored knowledge.

Theories that favor ecology, by contrast, tend to avoid using the notion of inference, and to understand perception as providing direct access to the world. Ecological approaches admit that perception involves, often complex, neurological activity. The brain has to ‘pick up’ the information present in the environment. However, according to ecologists, such activity is not inferential. The activity does not resemble reasoning, but simple attunement to the world.

Bayesian models are typically understood as belonging to the constructivist camp. Such models presume that the visual system forms and tests hypothesis in a way that resembles what scientists do.

In order to make sense of the notion of inference at the perceptual level – inferences that are fast, automatic, unconscious and that are not controlled by the subject – constructivists typically appeal to the idea of a series of mental representations that are manipulated in accordance with some encoded principles or assumptions. A process is inferential when, first, it occurs over states that, like the premises in an argument, have content. Perceptual hypotheses about half-spheres illuminated from above are examples of states that can serve as premises (and as conclusions).

In Bayesian models, good examples of premises are also the sensory states that inform about error, or that inform about the proximal stimulation.

Bayesian inferences are typically thought to occur in a multi-layered fashion where every level stands both for something in the world and for the pattern of stimulation that should be found at the level below.

Andy Clark, for example, says:

The representations units that communicate predictions downwards do indeed encode increasingly complex and more abstract features (capturing context and regularities at ever-larger spatial and temporal scales) in the processing levels furthest removed from the raw sensory input. In a very real sense then, much of the standard architecture of increasingly complex feature detection is here retained. (Clark 2012, p. 200)

Premises, however, are not the only elements of a mental inference. A transition is inferential if it follows a certain ‘logic’. Premises that merely succeed one another because of frequent co-occurrence, as in an association, do not manage to produce an inferential transition. What makes Bayesian inferences genuine inferences is the fact that they employ implicit knowledge or assumptions. Priors, as we saw in section 1, are thought to be encoded assumptions about what is more likely present in the environment.

Here is a small selection of quotes that confirms this point. Maloney and his collaborators say:

A Bayesian observer is one that has access to separate representations of gain, likelihood, and prior. (Maloney and Mamassian. 2009, p. 150)

Along similar lines, Gladzeijewski writes:

Since our system is supposed to realize Bayesian reasoning, there is one other aspect that the model structure and the environment structure should have in common. Namely, prior probabilities of worldly causes should be encoded in the generative model. (Gladzeijewski 2015, p. 14)

He adds:

What we need in order for this picture to be complete are specifics about the relevant structure of the representation itself. How exactly are the likelihoods, dynamics, and priors encoded or implemented in the nervous tissue? This is a vast and, to some degree, open subject. (Gladzeijewski 2015, p. 14)

Bayesian frameworks, then, commonly posit both the employment of premise-like representations and of represented principles. In this respect, they resemble other constructivist positions that also appeal to unconscious perceptual inferences.

In some versions of constructivism, the inference is ‘bottom-up’. This means that the inferential process starts from sensory representations of proximal conditions and it then uses assumptions as middle premises to derive percepts (Marr 1982, Rock 1983, Palmer 1999). The main difference with predictive coding theories is that the inference is driven from the bottom, rather than from high-level hypotheses. The basic idea, however is the same. In both cases, the visual system employs both sensory representations and represented assumptions.

Contrary to this way of categorizing Bayesianism, I want to now suggest that predictive coding accounts of perception are better seen as ecological approaches. This is because, if we exclude perceptual hypotheses, such accounts introduce structures that are not properly seen as representations. Bayesian inferences are not genuine inferences. They are biased processes that involve detectors.

Reflection on the notion of representation was once centered on giving naturalistically acceptable conditions for content, and for misrepresentation. More recently – partly due to the development of eliminativist projects in cognitive science, such as enactivism – the focus has somewhat shifted to give a *non-trivializing* characterization of what mental representations are. The worry is that naturalistic views of content based on function and/or on causal covariance, run the risk of promoting a notion of representation that is too liberal. They run the risk of confusing mere causal mediation – which simple detectors can do – with representation, which is a more distinctive psychological capacity (Ramsey 2007).

Detectors (or trackers) are typically understood as states that monitor proximal conditions and that cause something else to happen. As such, they are mere causal intermediaries. What is important about detectors is that they are active in a restricted domain, and they do not model distal or absent conditions. The behavior of a system that uses only detectors is mostly viewable as controlled by the world. What the system does can be explained by appeal to present environmental situations which detectors simply track. Magnetosomes in bacteria and cells in the human immune system that respond to damage to the skin are examples of detectors of this kind.

The development of eliminativist views that recommend the elimination of representations from scientific accounts of the mind, prompted the shift in focus from content to the representation-detection distinction because a notion of representation that is too liberal prevents a meaningful dispute (Chemero 2009). No eliminativist denies that there are trackers – that is, that there is causal, lawful mediation between environmental stimuli and behavior. What eliminativists rather deny is that there is representational mediation between the two.

Representations are typically taken to be internal, unobservable structures of an organism that carry information about conditions internal or external to the organism itself. Representations can do so by having different formats. A map-like representation carries information by having a sort of structural similarity or isomorphism to its subject matter. A pattern of spatial relations, for example, is preserved between a map and the terrain it represents. A pictorial representation may employ mere resemblance, while a linguistic representation employs neither resemblance or isomorphism.

Other fairly uncontroversial features of representations are that they guide action and that they can misrepresent. Maps guide us through a city, pictures can tell us what to expect in a given environment. Action in this context is not limited to what involves bodily movement. It can consist in mental action – such as mental inference. Representations guide action and can misrepresent in the sense of giving us false information. Bad maps and pictures exemplify this capacity for error.

In addition to these basic conditions, an equally central feature of representation that has received some sustained attention is the idea that representations are ‘detachable’ structures (Clark and Toribio 1994, Gladziejewski 2015, Grush 1997, Ramsey 2007). Mere causal mediators track what is present. Representations, by contrast, display some independence from what is around.

There are different ways of spelling out what the detachment of representations amounts to. Different proposals agree on the common theme that representations are employed when an organism’s behavior is not controlled by the world, but rather by the content of an internal model of the world. In some theories, this idea is understood in terms of ‘off-line’ use. Representations can guide action even when what is represented by them is not present.

This kind of offline detachability does not have to be very radical. A GPS, for example, is a representational device that guides action while being

coupled with the city it represents. The GPS is an example of a representation because one's ongoing decisions concerning where to go are dictated by the GPS. They depend on the contents of the GPS that serves as a stand-in. The decisions are not directly controlled by the world (Gladziejewski 2015, p.10).

In other proposals, the detachment of representations is spelled out in terms of their content. Representations allow us to coordinate with what is absent because they stand for what is not present to the sensory organs (Orlandi 2014). A perceptual state that stands for a full object even when the object is partly occluded or out of view, for example, counts as a representation. It allows perceivers to coordinate with what is not, strictly speaking, present to their senses.

Regardless of how the idea of detachability is fleshed out, in what follows we can use the common understanding of representations as unobservable structures that are capable of misrepresenting, that guide action and that are detachable in one of the senses just described. In using this notion we are not departing from common usage. We are using a notion that is explicitly accepted by proponents of Bayesian accounts of perception (Clark and Toribio 1994, Gladziejewski 2015).

My claim is that predictive coding models introduce structures that are better seen as non-representational. Low and intermediate-level hypotheses, and sensory states are better seen as mere causal mediators. Priors, hyperpriors and likelihoods are better seen as mere biases.

Before proceeding to explain why this is the case, it may be useful to point out that the success of Bayesian theory does not commit one to posit processes of inference. The fact that a system acts *as if* it is reasoning does not imply that it is.

This is confirmed by the fact that Bayesian models have been offered for a variety of phenomena some of which plausibly involve representational and psychological states, and some of which do not. We have bayesian accounts of color constancy, decision-making and of action generation. We also have, however, bayesian models of the purely physiological activity of retinal ganglion cells in salamanders and rabbits (Hosoya et al. 2005, p. 71). Such activity involves no representations. Retinal cells can act *as if* they are testing hypotheses without genuinely doing so.

Indeed, this is true even if one has fairly robust realist tendencies. One may think that if a model is very successful in predicting what something does, then it must map onto something real. If Bayesian accounts are very

good at explaining diachronic aspects of perception – for example, why a percept is retained in the face of variations in incoming stimulation – then the accounts must describe some deep reality of the perceptual system.

Granting this point, however, does not yet commit us to the nature of what is real. All it tells us is that we should find features that make the perceptual system perform in a way that the model successfully describes. One can accept, for example, the description of the brain as reducing error, while not yet committing to the nature of what exactly does the error reduction – whether it involves representations or not. Error reduction can be spelled out, for example, in terms of free-energy minimization, a characterization that does not make reference to psychological states (Clark, 2012 p. 17, Friston and Stephan 2007, Friston 2010).

The question is then whether predictive coding models posit representations. Given our preliminary portrayal, the perceptual hypotheses that are formed in response to the driving signal are plausibly representational states. They are states that stand, for example, for convex half-spheres illuminated from above. They guide action, they can misrepresent and, depending on how we understand detachability, they have some sort of independence from what is present in the environment. In the case of figure 1 they stand for something absent, since there is nothing convex in the two-dimensional image. Perhaps, there is some room to argue that such states are not usable off-line, but I think that we can gloss over this issue and admit that the deliverances of visual processing are representational states. In this sense, Bayesian accounts plausibly introduce representational structures.

It is a further question, however, whether the states that precede high-level perceptual hypotheses are themselves representations. I think that they are not.

To start, we should point out that there is an already noticed tension in how Bayesians talk of levels of perception that precede high-level hypotheses. Sensory states seem to have a triple function in predictive coding accounts. They are supposed to carry information about what is in the environment, in particular about proximal conditions, such as the characteristics of light reflected by objects. They also have to serve as error signals telling the brain whether the current hypothesis is plausible. Finally, a level of sensory states serves as mock input. This is the level produced by the current winning hypothesis for purposes of testing.

Leaving aside the issue of which of these functions is more prominent, we can point out that sensory states, in each of these conceptions, either do not

guide action, or are not detached to count as representations (or both). If we conceive of them, first, as informants of proximal conditions, such states are not properly detached from their causal antecedents and they do not steer what the subject does. Areas of the retina, the optic nerve or the primary visual cortex, that carry information about light patterns, gradients and other elements of the proximal input do not guide person-level activity, and are not properly detached from the world. They merely track what is present and activate higher areas of perceptual processing.

Although, in Bayesian frameworks, these low and intermediate stages of processing are called ‘hypotheses’ this only means that they predict what is present at the level below them. It does not mean that the stages, like their perceptual high-level counterparts, are about well-defined environmental quantities and drive action. It is in fact an insight of predictive coding models – and something that distinguishes them from more traditional bottom-up constructivist models – that early and intermediate visual states do not track well-defined elements. Neurons in early visual areas do not have fixed receptive fields, but respond in some way to every input to which they are exposed (Jehee and Ballard 2009, Rao and Sejnowski 2002). Sensory states of this kind are mere detectors or mediators.

We reach a similar conclusion, if we regard early visual stages as error detectors. The initial tendency may be to suppose that the error signal is a well-determined and informative signal – telling the brain what aspects of the current hypothesis are wrong. This, however, is both not mandatory and not plausible. There is little evidence for an independent channel of error signals in the brain (Clark 2012). Low-level perceptual stages appear to communicate the kind of error present by gradually attempting to reconfigure the hypothesis entertained. They simply provide ongoing feedback by providing ongoing information about proximal conditions.

Error signals understood in this way are both not driving action – unless we question-beggingly assume, as Gladziejewsk (2015) seems to do, that action *is* error reduction – and not sufficiently detached from present conditions. Indeed, in a sense, the signals are not even *about* the conditions present to the perceiver. If we look at them as *just* error signals, then they are about things internal to the brain. They inform the brain concerning the need to adjust its own states to reach an error-free equilibrium. They are ‘narcissistic’ detectors in a similar sense to the one outlined by Akins (1996).

Mock sensory states are similarly narcissistic. They are presumably produced by the winning high-level hypothesis to run a simulation of what the

incoming data would have to be like if the hypothesis were true. Admittedly, this is an idealized way of describing perceptual activity. But even conceding this description, such states only concern the brain's internal states. Their aim is the brain's equilibrium. They are not aimed at steering action and at doing so in the absence of the sensory stimulation they are supposed to match. They are states produced *for* checking the level below them. This exhausts their function.

If this is true, then we have reasons to doubt that low and intermediate sensory states are representations. In Bayesian accounts, such states do not have the function of mapping the distal environment and guiding activity. In fact, in contrast to more traditional bottom-up accounts, the metaphor of a construction is strange when applied to predictive coding frameworks. In predictive coding, percepts are not constructed or built from the ground up. They are rather triggered by – and then checked against – incoming stimulation.

Consider, next, the presumed encoded assumptions of Bayesian models. Priors, hyperpriors and likelihoods are presumed to be encoded pieces of knowledge that make perception a genuine inferential process. They are used to prefer and discard certain hypotheses in the face of messy data. It is hard to see, however, why we should see priors, hyperpriors and likelihoods in this over-intellectualized way. An alternative – entertained and discarded by some proponents of predictive coding accounts – is that priors, hyperpriors and likelihoods are mere biases (Hohwy 2013).

A bias is a functional, non representational feature of a system that can be implemented in different ways and that skews a system to perform certain transitions or certain operations. In implicit bias, for example, the brain presumably has features that skew it to move from certain concepts to certain other concepts in an automatic fashion due to frequent concomitant exposure to the referents of the concepts. One may automatically move from the concept QUEER to the concept DANGEROUS as a result of being exposed to many images of dangerous queers (or of the danger of certain alleged queer practices).⁸

In a biological brain, a bias can be given by an innate or developed axonal connection between populations of neurons. The connection is such that the populations are attuned to each other. The activity of one increases the

⁸It is in fact a matter of dispute whether implicit biases are underwritten by simple biases of this kind (Mandelbaum 2015). For reasons of space, I leave this dispute aside.

probability of the activity of the other.

Understanding perceptual priors, hyperpriors and likelihoods as biases means thinking that, as a result of repeated exposure to the structure of the world in the evolutionary past *and* in the present, the perceptual system is skewed to treat certain stimuli in a certain way. It is skewed, for example, to treat certain stimuli as coming from convex objects, or to treat certain discontinuities in light as edges. It is further skewed to continue to treat the stimulus in this set way unless a substantive change in stimulation happens.

We can of course refer to biases as encoded assumptions, but that is because pretty much everything that acts to regiment a system can be called an assumption without being one. Biases may be present in the differential responses of plants to light sources, in the training of pigeons to pick certain visual stimuli, in the retractive behavior of sea slugs and in many other systems that can only metaphorically be described as having implicit knowledge or assumptions (Frank and Greenberg 1994).

A cluster of converging evidence points to the fact that we should treat priors, hyperpriors and likelihoods as mere biases. First, in predictive coding accounts, priors, hyperpriors and likelihoods do not look like contentful states of any kind. They are not map-like, pictorial or linguistic representations. Having accuracy conditions and being able to misrepresent are not their central characteristics. Priors, hyperpriors and likelihoods rather look like built-in or evolved functional features of perception that incline visual systems toward certain configurations.

To bring out this point, we can again contrast Bayesian perceptual inferences to more traditional – bottom-up – perceptual inferences. In bottom-up inferences sensory states are premises, and encoded assumptions serve as additional premises in drawing conclusions about what is in the world. Priors and likelihoods, however, do not serve as premises for visual inferences. They are not independent structures that enter into semantic transitions with other premises. They rather have the simple function of marking a hypothesis as more or less probable. They are like valves. They skew the brain toward certain neuronal arrangements.

Second, typically (although not invariably) encoded assumptions, unlike biases, are available in a variety of mental contexts and they may show up in the subject's conscious knowledge. They may guide mental or bodily action. Priors, hyperpriors and likelihoods, however, are proprietary to the perceptual apparatus. They do not show up in different contexts. They are proprietary responses to the environment we inhabit.

In describing priors and hyperpriors, I simplified their formulation for expository purposes. The prior concerning light coming from overhead, for example, has it that light comes from above *and slightly to the left*, a fact that is yet to be explained. Subjects are not aware of this prior, or of any of the others that regulate their perceptual achievements.

Third, a recognized sign that something is a biased process is that it is not sensitive to contrary evidence. If you are biased to think that queers are dangerous, then reasoning will do little to disrupt this bias. Biases are typically only sensitive to environmentally-based interventions, such as extinction or counter conditioning.

Now, perception is famously stubborn to contrary evidence, as perceptual illusions amply demonstrate (Kanizsa1985, Gerbino and Zabai 2003). Even the controversial phenomenon of cognitive penetrability can be understood in the framework of stubborn perceptual systems exposed to various environmental regularities, and driven by attentional mechanisms (Orlandi 2014, Pylyshyn 1999).

The moral of this section is that Bayesian accounts have features that support a non-constructivist interpretation. The machinery they posit is easily seen as non-representational, save for the hypotheses that end up being percepts. Indeed, such accounts seem to just offer an ecologically plausible addition to Natural Scene Statistics. They offer a description of how the brain attunes to the structure of the environment in which it is situated. Since this may be surprising from a historical point of view, the next section hopes to make this moral more plausible.

5 Brief historical aside

If what I argued so far is right, then we should consider re-categorizing Bayesian accounts of perception as ecological and Gibsonian accounts. Bayesian perceptual theory needs to be grounded in Natural Scene Statistics and its posits are better understood as descriptions of how the brain attunes to the environment it inhabits.

This is surprising for a couple of reasons. The first reason is that proponents of predictive coding models identify Helmholtz as their predecessor, and Gibson was largely reacting against Helmholtz's work when he proposed his ecological ideas. The second reason is that Gibson himself would not be happy with admitting that percepts are representational states – something

that I granted in the last section. This aspect of predictive coding models would not be congenial to an ecologist.

Now, with respect to the second point, if we read Gibson as an eliminativist, then we should expect him to have a generalized disdain for representations. Eliminativists tend to want to avoid this notion altogether.

I follow other commentators, however, in supposing that Gibson's main target was not primarily the idea of percepts as representations, but the idea that perception involves an inference (Rowlands 1999). This idea was introduced at the end of the 1800 by Helmholtz who was in turn influenced by the work of Kant. Helmholtz initiated a constructivist tradition that includes Marr, Gregory, Rock and Palmer.

Theorists in this tradition do not merely claim that the deliverances of perceptual systems are representations. They also claim that such deliverances are the product of an inferential process. Perceptions are *mediated* responses to the environment.

Gibson reacted to this type of position. In his view, perception is direct because it is determined by environmental regularities. In a line of reasoning similar to the one employed in section 2, Gibson claimed that appeal to the richness of environmental information – what he called ‘invariances’ – makes recourse to internal inferences superfluous.⁹

Interestingly, Gibson understood that the brain needs to attune and respond to the rich environment it encounters. He never claimed that perceiving is mechanistically simple. He agreed, for example, that we need internal states that pick-up information. He failed, however, to detail how the perceptual mechanism works, often using, instead, a resonance metaphor. According to him, the information in visual stimuli simply causes the appropriate neural structures in the brain to fire and resonate (Gibson 1966, 1979). This resonance metaphor remained underdescribed in Gibson, and it was later at the center of criticism (Marr 1982).

It turns out, however, that this metaphor can be spelled out precisely in predictive coding terms. In formulating the resonance metaphor, Gibson was inspired by the work of Gestalt Psychologist Wolfgang Köhler, in particular by Köhler's idea that the brain is a dynamic, physical system that converges towards an equilibrium of minimum energy. This idea predates contemporary

⁹Action also plays a central role in Gibson's view. For reasons of space, I leave the topic of action in perception undiscussed. It is sufficient to say that this aspect of Gibson's approach is in line with predictive coding models where action is one of the ways in which the brain may react to the inadequacy of a hypothesis.

accounts of the brain as Bayesian systems (Clark 2012, Köhler 1920/1950).

The development of connectionist and artificial neural networks later put ‘flesh and bones’ on Köhler’s idea, and such networks are the model of choice for predictive coding frameworks as well (Churchland 1990, Feldman 1981, Rumelhart and McClelland 1988, Rosenblatt 1961).

Artificial neural networks are nets of connected units that spread information by spreading levels of activation, simulating a biological brain. There are different ways of building networks, but some can learn to attune to the environment by using simple units and connections, rather than more sophisticated symbolic and representational structures. They can learn, for example, to detect the presence of faces in a presented scene.

In networks of this kind, learning is due primarily to two factors. One is repeated exposure to environmental stimuli, or to the ‘training set’, which may consist in a set of images of a wide variety of faces. The other factor is the network’s ongoing attempt to *reduce error*. The reduction is describable in Bayesian terms, and it consists primarily in adjusting the connection strengths between units so that the network responds better to the stimulation. Repeating the process of weight adjustment often enough, makes it so that networks develop internal biases that skew them to respond in the right way – signaling the presence of faces –whenever a certain kind of proximal stimulation is present. Networks of this kind continuously attempt to reach a state of internal equilibrium and, by so doing, they tune to what is in the world.

There is ongoing disagreement concerning whether connectionist networks make use of representations *at all*. Compared to classical, artificial models of intelligence, they are more open to non-representational interpretations (Ramsey 1997, Ramsey et al. 1991). This is true, not just because such models employ no symbols and explicit algorithms, but also because they do not seem to employ map-like or pictorial representations either. Their processing units, particularly at the lower and intermediate levels, can be seen as mere causal intermediaries. The connections between the units can be seen as mere biases. We can understand the kind of processing that networks execute without appeal to representations.

The overall historical point is that there is a certain continuity between Gibson’s resonance metaphor and the idea that perceptual systems are error reducers modeled by connectionist networks. There is, on the other hand, a discontinuity between predictive coding accounts of perception and traditional bottom-up inferential accounts of the kind proposed by Marr and Rock.

Such accounts are usually supported by reference to classical computational models (Fodor and Pylyshyn 1981, Ullman 1979).

It is then not clear why Bayesians should think of themselves as belonging to the inferential tradition. Cognitive science has presumably moved away from the representation-heavy computations of classicism. Bayesian perceptual theory is a recent development better modelled in connectionist nets. It is then plausibly an ecologically-friendly addition to the study of environmental contingencies, rather than a constructivist alternative.

6 Conclusion

In this article, I argued for two main claims. One is the claim that Bayesian accounts, as they are typically described, are not well-suited to explain how the inverse problem is solved. Such accounts need to admit the larger role of the driving signal, and be integrated with an ecological study of the environmental regularities in which vision occurs. They need to be integrated with a developing branch of perceptual psychology called Natural Scene Statistics.

I argued further that Bayesianism is in fact a fitting addition to NSS because, with the exception of perceptual hypotheses, it postulates structures that are not properly seen as representations. Bayesian perceptual inferences are not genuine inferences. They are biased processes that operate over detectors. If both of these claims are true, then Bayesian perception is not constructivist perception. It is ecological perception.

References

- Adams, W. J., Graf, E. W., and Ernst, M. O. (2004). Experience can change the 'light-from-above' prior. *Nature neuroscience*, 7(10):1057–1058.
- Akins, K. (1996). Of sensory systems and the "aboutness" of mental states. *The Journal of Philosophy*, 93(7):337–372.
- Beierholm, U. R., Quartz, S. R., and Shams, L. (2009). Bayesian priors are encoded independently from likelihoods in human multisensory perception. *Journal of vision*, 9(5):23.

- Brainard, D. H. (2009). Bayesian approaches to color vision. *The visual neurosciences*, 4. available online at <http://color.psych.upenn.edu/brainard/papers/BayesColorReview.pdf>.
- Chemero, A. (2009). *Radical embodied cognitive science*. The MIT Press.
- Churchland, P. (1990). Cognitive activity in artificial neural networks. In Osherson, D. and Smith, E., editors, *Thinking: An invitation to cognitive science*. MIT Press, Cambridge, MA.
- Clark, A. (2012). Whatever next? predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*.
- Clark, A. and Toribio, J. (1994). Doing without representing? *Synthese*, 101:401–431.
- Feldman, J. (1981). A connectionist model of visual memory. *Parallel Models of Associative Memory: Updated Edition*, page 65.
- Fodor, J. A. and Pylyshyn, Z. W. (1981). How direct is visual perception? some reflections on gibson’s” ecological approach.”. *Cognition*.
- Frank, D. A. and Greenberg, M. E. (1994). Creb: a mediator of long-term memory from mollusks to mammals. *Cell*, 79(1):5.
- Friston, K. (2010). The free-energy principle: a unified brain theory? *Nature Reviews Neuroscience*, 11(2):127–138.
- Friston, K. J. and Stephan, K. E. (2007). Free-energy and the brain. *Synthese*, 159(3):417–458.
- Geisler, W. (2008). Visual perception and the statistical properties of natural scenes. *Annu. Rev. Psychol.*, 59:167–192.
- Geisler, W., Perry, J., Super, B., Gallogly, D., et al. (2001). Edge co-occurrence in natural images predicts contour grouping performance. *Vision research*, 41(6):711–724.
- Gerbino, W. and Zabai, C. (2003). The joint. *Acta Psychologica*, 114:331–353.

- Gibson, J. (1966). *The senses considered as perceptual systems*. Houghton Mifflin, Boston.
- Gibson, J. (1979). *The Ecological Approach to Visual Perception*. Houghton Mifflin, Boston.
- Gładziejewski, P. (2015). Predictive coding and representationalism. *Synthese*, pages 1–24.
- Gorman, R. P. and Sejnowski, T. J. (1988). Analysis of hidden units in a layered network trained to classify sonar targets. *Neural networks*, 1(1):75–89.
- Gregory, R. (1966). *The Intelligent Eye*. McGraw-Hill, New York.
- Grush, R. (1997). The architecture of representation. *Philosophical Psychology*, 10(1):5–23.
- Helmholtz von, H. (1867/1925). *Treatise on physiological optics*, volume 3. Courier Dover Publications.
- Hohwy, J. (2013). *The predictive mind*. Oxford University Press.
- Hohwy, J., Roepstorff, A., and Friston, K. (2008). Predictive coding explains binocular rivalry: an epistemological review. *Cognition*, 108(3):687–701.
- Hosoya, T., Baccus, S. A., and Meister, M. (2005). Dynamic predictive coding by the retina. *Nature*, 436(7047):71–77.
- Howe, C. Q., Beau Lotto, R., and Purves, D. (2006). Comparison of bayesian and empirical ranking approaches to visual perception. *Journal of Theoretical Biology*, 241(4):866–875.
- Jehee, J. F. and Ballard, D. H. (2009). Predictive feedback can account for biphasic responses in the lateral geniculate nucleus. *PLoS computational biology*, 5(5):e1000373.
- Kanizsa, G. (1985). Seeing and thinking. *Acta Psychologica*, 59(1):23–33.
- Köhler, W. (1920/1950). Physical gestalten. pages 17–54.

- Maloney, L. T., Mamassian, P., et al. (2009). Bayesian decision theory as a model of human visual perception: testing bayesian transfer. *Visual neuroscience*, 26(01):147–155.
- Mamassian, P. and Landy, M. S. (1998). Observer biases in the 3d interpretation of line drawings. *Vision research*, 38(18):2817–2832.
- Mamassian, P., Landy, M. S., Maloney, L. T., Rao, R., Olshausen, B., and Lewicki, M. (2002). Bayesian modelling of visual perception. *Probabilistic models of the brain: Perception and neural function*, pages 13–36.
- Mandelbaum, E. (2015). Attitude, inference, association: On the propositional structure of implicit bias. *Nous*.
- Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*, henry holt and co. Inc., New York, NY.
- Marr, D. and Hildreth, E. (1980). Theory of edge detection. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, 207(1167):187–217.
- Obermayer, K., Sejnowski, T., and Blasdel, G. (1995). Neural pattern formation via a competitive hebbian mechanism. *Behavioural brain research*, 66(1):161–167.
- Orlandi, N. (2014). *The innocent eye: why vision is not a cognitive process*. Oxford University Press.
- Palmer, S. E. (1999). *Vision science: Photons to phenomenology*, volume 1. MIT press Cambridge, MA.
- Purves, D., Wiedner, M. A., and Lotto, R. B. (2003). *Why we see what we do: An empirical theory of vision*. Sinauer.
- Pylyshyn, Z. et al. (1999). Is vision continuous with cognition? the case for cognitive impenetrability of visual perception. *Behavioral and brain sciences*, 22(3):341–365.
- Ramachandran, V. S. (1988). Perceiving shape from shading. *Scientific American*, 259(2):76–83.

- Ramsey, W. (1997). Do connectionist representations earn their explanatory keep? *Mind & language*, 12(1):34–66.
- Ramsey, W., Stich, S., and Garon, J. (1991). Connectionism, eliminativism, and the future of folk psychology. In *Philosophy and connectionist theory*, pages 199–228. Hillsdale, NJ: Lawrence Erlbaum.
- Ramsey, W. M. (2007). *Representation reconsidered*. Cambridge University Press.
- Rao, R. P. and Sejnowski, T. J. (2002). Predictive coding, cortical feedback, and spike-timing dependent plasticity. *Probabilistic models of the brain: Perception and neural function*, page 297.
- Rescorla, M. (2013). Bayesian perceptual psychology. In Matthen, M., editor, *The Oxford Handbook of the Philosophy of Perception*. Oxford University Press.
- Rescorla, M. (2015). Review of nico orlandi’s the innocent eye. *Notre Dame Philosophical Reviews*.
- Rescorla, M. (forthcoming). Bayesian sensorimotor psychology. *Mind and Language*.
- Rock, I. (1983). *The logic of perception*. MIT press Cambridge, MA.
- Rosenblatt, F. (1961). Principles of neurodynamics. perceptrons and the theory of brain mechanisms. Technical report, DTIC Document.
- Rowlands, M. (1999). *The body in mind: Understanding cognitive processes*. Cambridge University Press.
- Rumelhart, D. E., McClelland, J. L., Group, P. R., et al. (1988). *Parallel distributed processing*, volume 1. IEEE.
- Sejnowski, T. J. and Rosenberg, C. R. (1987). Parallel networks that learn to pronounce english text. *Complex systems*, 1(1):145–168.
- Shi, Y. Q. and Sun, H. (1999). *Image and video compression for multimedia engineering: Fundamentals, algorithms, and standards*. CRC press.

- Stone, J. V. (2011). Footprints sticking out of the sand (part ii): Children's bayesian priors for shape and lighting direction. *Perception*, 40(2):175–190.
- Ullman, S. (1979). *The interpretation of visual motion*, volume 28. MIT press Cambridge, MA.
- Yang, Z. and Purves, D. (2004). The statistical structure of natural light patterns determines perceived light intensity. *Proceedings of the National Academy of Sciences of the United States of America*, 101(23):8745–8750.