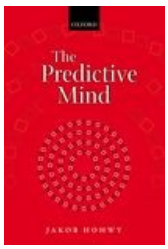

The Predictive Mind

Jakob Hohwy

https://doi.org/10.1093/acprof:oso/9780199682737.001.0001

Published: 2013 **Online ISBN:** 9780191766350 **Print ISBN:** 9780199682737

Search in this book

CHAPTER

10 Perceptual unity in action

Jakob Hohwy

https://doi.org/10.1093/acprof:oso/9780199682737.003.0011 Pages 207–223

Published: November 2013

Abstract

This chapter explores the link between prediction error minimization and conscious perception that the previous chapter put on the table. It then goes on to discuss a special property of conscious perception, namely its unity. This is an important but evasive part of our phenomenology. To explain it one needs to deal with at least two aspects: (1) why the elements of conscious unity are *united*; (2) why conscious unity is *unitary*. The answer to this issue begins by setting out a key theory consciousness, namely the global neuronal workspace theory, and by noting how its key notion of 'ignition' is in need of explanation. By appealing to prediction error minimization, and in particular the notion of active inference, an interesting explanation of ignition, and in turn of perceptual unity becomes available. The chapter ends with a more general discussion of the implications of this appeal to action for our conception of the mind in the world.

Keywords: unity of consciousness, active inference, global neuronal workspace theory, ignition, mind-world relation

Subject: Philosophy of Perception, Philosophy of Science, Philosophy of Mind

Collection: Oxford Scholarship Online

This chapter explores the link between prediction error minimization and conscious perception that the previous chapter put on the table. It draws out the reasons why there might be such a link and why it might be interesting. Then I discuss a special property of conscious perception, namely its unity. This is an important but evasive part of our phenomenology. To explain it one needs to deal with at least two aspects: (1) why the elements of conscious unity are *united*, that is how they hang together; this is an issue that harks back to the binding problem of Chapter 5, but which here is given a more general treatment; (2) why conscious unity is *unitary*, that is why there is only one overall perceptual state which subsumes all other states, and not more. By appealing to prediction error minimization, and in particular the notion of active inference, an interesting explanation of these two aspects of perceptual unity becomes available.

Downloaded from https://academic.oup.com/book/4105/chapter/145830336 by Royal Library Aarhus University user on 19 April 2024

From causal inference to consciousness?

Perception is the business of inferring the causes of our sensory input. This feat is performed by a prediction error minimization mechanism that is replicated hierarchically up through the brain. Prediction error minimization in hierarchical inference is able to capture much of our first-person perspective; it makes room for perceptual binding of sensory attributes, as well as the ways in which prior beliefs can penetrate perceptual experience, and the ways in which perception can be disrupted. It provides a framework for understanding attention and how attention relates to conscious experience. At least these are the things I have been arguing in the foregoing chapters. I have also made the observation, in Chapter 8, that these aspects of perception, which the prediction error mechanism can accommodate, feature high on the list of canonical aspects of *conscious* perception. It is tempting then to pursue further the idea that whereas perceptual inference itself is unconscious, the upshot of it is conscious experience.

p. 208 I find this a good position to be in, for someone interested in the nature of consciousness. Rather than setting out with the intention to reveal the nature of consciousness we are looking at a mechanism that solves the problem of perception by appeal to internal generative models, and we are discovering that this mechanism may harbour just the aspects we consider characteristic of consciousness. This type of idea is captured nicely by Thomas Metzinger:

> [A] fruitful way of looking at the human brain [...] is as a system which, even in ordinary waking states, constantly hallucinates at the world, as a system that constantly lets its internal autonomous simulational dynamics collide with the ongoing flow of sensory input, vigorously dreaming at the world and thereby generating the content of phenomenal [i.e., conscious] experience. (Metzinger 2004: 52)

This kind of explanatory potential is part of the reason why the prediction error minimization idea is so attractive. But care is needed in how the question of consciousness is handled. If the question is the straightforward one whether prediction error minimization is the mechanism in the brain that explains why some perceptual content is conscious rather than not conscious, then the answer is 'no'. To explain this answer we need to recall the disclaimer from the previous chapter that there is little reason to think that we can solve the hard problem of consciousness in its most thorny version (Chalmers 1996; Levine 2001).

But there is something else we can do to answer the question whether prediction error minimization is the mechanism in the brain that explains why some perceptual content is conscious rather than not conscious. We will not explain why there is *something it is like* for the creature, but we can answer it by focusing on *properties* of conscious perception. The use of 'conscious' in expressions like "...why some perceptual content is conscious rather than not conscious..." should be taken to mean 'conscious, that is, perceptual inference that is characterized by first-person perspective, inextricable richness, sensory binding, is cognitively penetrable, is subject to illusions and reality testing, relates in an intimate way with attention etc.'.

This conception of conscious perception thus consists of a core cluster of descriptors arising from conceptual analysis of our common-sense concept of conscious perception, and this is combined with relatively broad empirical information about enduring projects in cognitive psychology (finding the binding mechanism), psychophysics (explaining illusions), and psychopathology (mental illness). The idea is that we should explore the underlying mechanism that brings about these characteristics because it is close enough to capturing many of the things we care about in conscious experience. If we could find a mechanism that explains all of these characteristics, then we will have made inroads on the nature of consciousness.

Downloaded from https://academic.oup.com/book/4105/chapter/145830336 by Royal Library Aarhus University user on 19 April 2024

p. 209 There is one characteristic of conscious experience that I have not discussed in any of the preceding chapters. This is the unity of conscious perception. There is also one aspect of prediction error minimization that I have not yet brought to bear much on the characteristics of conscious perception. This is active inference. This chapter will attempt to remedy this, because I think active inference is the key to understanding the unity of conscious perception.

Perceptual unity

When you introspect you find that your mind is populated by conscious experiences. There is a constant flow of sensations, perceptions, moods, thoughts. One striking feature is that at any one time all these conscious aspects are bound together in one unified conscious field. We never seem to find two or more completely distinct conscious streams when we introspect. This is phenomenal unity ('phenomenal' means 'conscious' but picks out a characteristic that pertains specifically to the qualitative aspect of consciousness rather than other notions of unity relating to personal identity, or mere representational unity; I shall focus on the subset of these states that are perceptual).

Tim Bayne argues convincingly that consciousness is always phenomenally unified. Here is how he explains the notion:

> Let us say that a subject has a unified consciousness if, and only if, every one of their conscious states at the time in question is phenomenally unified with every other conscious state. We can think of such subjects as fully unified. Where a subject is fully unified, we can say that they enjoy a single total conscious state. A total conscious state is a state that is subsumed by nothing but itself, where one conscious state subsumes another if the former includes the latter as a 'part' or 'component'. This total state will capture what it is like to be the subject at the time in question. In specifying the subject's total conscious state one thereby provides a full specification of the subject's specific conscious states. By contrast, if we are dealing with a creature whose consciousness is disunified, then there will be no single conscious state that 'subsumes' each of their specific conscious states. (Bayne 2010: 15)

Bayne argues that consciousness is unified in that sense:

> How common is it for human beings to have a unified consciousness in this sense of the term? Quite common, it seems to me. Take any set of conscious states that you are currently enjoying—visual experiences, auditory experiences, emotional experiences, bodily sensations, conscious thoughts, or whatever. Irrespective of the degree to which these states might be representationally unified with each other, they will—I wager—be mutually phenomenally unified with each other.
> p. 210 You might not have been aware of this unity until I drew your attention to it, but ↳ having drawn your attention to it I trust that you recognize it in your own experience. (ibid.)

And he uses this to propose that unity is essential to consciousness:

> Generalizing somewhat, we might hazard the guess that unity in this sense is a deep feature of normal waking experience. Indeed, we might even go further, and suggest that this kind of unity is not just a feature of normal waking experience but also characterizes other kinds of background states of consciousness, such as those that are present in REM dreaming, hypnosis, and various pathologies of consciousness. One might even hazard the thought that this kind of unity is an essential feature of consciousness—one that it cannot lose (at least when it comes to creatures like us). (ibid.)

Downloaded from https://academic.oup.com/book/4105/chapter/145830336 by Royal Library Aarhus University user on 19 April 2024

This is the *unity thesis*, that "the conscious states that any subject of experience enjoys at any one point in time will occur as the components of a single total phenomenal state—a unitary 'phenomenal field' (Bayne 2010: 75). Bayne's strategy is to present the thesis, appeal to introspection to defend it (and defend this against objections), and then show how purported cases of disunity fail to be persuasive or to target efficiently the unity thesis.

I think the unity thesis as stated and defended by Bayne is true and captures an essential characteristic of our consciousness. I will consider how and to what extent the prediction error minimization mechanism can help us understand phenomenal unity. That is, why should a prediction error minimizing system like us have unified rather than disunified conscious states?

Before starting on this task, some preliminaries. First a limitation. Bayne discusses unity for all conscious experience; I will only consider conscious *perception*. This is because we have so far mainly considered perceptual inference and been silent on the extent to which other conscious experiences (moods, emotions, thoughts) can be explained by prediction error (for an attempt at emotion, see Hohwy 2011; Seth, Suzuki et al. 2012; and Chapter 12). For this reason, I will discuss a more limited notion of what I call *perceptual unity*. Transposed to Bayne's idiom this is the idea that the conscious perceptual states that any subject of experience enjoys at any one time will occur as the components of a single subsuming total phenomenal perceptual state.

The second preliminary is that the unity thesis, even though it is offered as something essential about consciousness, is instead a contingent truth about our consciousness. There is no conceptual pressure towards unity—mere a priori, conceptual analysis of the concept of consciousness does not lead to the unity thesis. It might be difficult to positively conceive what it would be like to have a genuinely disunified conscious state but it doesn't seem impossible in the sense square circles are impossible. For this reason, the mechanism that guarantees unity for us can be quite specific to our neuronal machinery and allow that creatures in a relatively close possible world to us can fail to have unified consciousness.

p. 211 This is an important facet, when trying to locate unity. It means that we can look for a genuinely causal property that we happen to possess, rather than something any conscious creature must have irrespective of which logically possible world it inhabits. It also means that we can attribute perceptual unity to ourselves but can withhold it to other creatures even if we allow the possibility that they are conscious. Finally, contingency rather than necessity comes with more of a risk of fragility. Even though Bayne is right and we are all in fact always enjoying perceptually unified experience it might still be that, at least in principle, it wouldn't take much to upset the phenomenal applecart and produce disunified perception. The trick would be to explain why this doesn't in fact ever seem to happen to us.

Unity, and ignition of the global neuronal workspace

It would count strongly in favour of a theory of consciousness if it could explain perceptual unity. Theorizing on consciousness is still somewhat in its infancy but among the more developed theories Bayne singles out Tononi and Edelman's dynamic core theory as the one most likely to accommodate unity (Edelman and Tononi 2000). As Bayne points out, this is no surprise since this theory is one of the few that begins with unity as a constraint on the theory (Bayne 2010: Ch. 10). According to this intriguing theory, thalamocortical loops (i.e., looping neuronal pathways between the cortex and the deeper thalamic structures in the brain) sustain the conscious field in which perceptual content is unified.

I want to make one further move with respect to theories of consciousness. The basic idea motivating the dynamic core theory is included in the kind of *global neuronal workspace theory*, which Stanislas Dehaene, Jean-Pierre Changeux, and Lionel Naccache and their colleagues are developing. Dynamic core theory builds

Downloaded from https://academic.oup.com/book/4105/chapter/145830336 by Royal Library Aarhus University user on 19 April 2024

on the notion of informational re-entry, which is the bidirectional exchange of signals across parallel cortical maps coding for different aspects of the same object. This is what helps build up a unified percept, according to this theory. Dehaene and colleagues sought to integrate this idea with a wider range of characteristics of consciousness such as executive supervision and limited capacity, and to that end introduced the global neuronal workspace theory:

> We [propose] the Global Neuronal Workspace (GNW) model as an alternative cortical mechanism capable of integrating the supervision, limited-capacity, and re-entry properties. [...] Global broadcasting allows information to be more efficiently processed (because it is no longer confined
p. 212 to a subset of nonconscious ↳ circuits but can be flexibly shared by many cortical processors) and to be verbally reported (because these processors include those involved in formulating verbal messages). Nonconscious stimuli can be quickly and efficiently processed along automatized or preinstructed processing routes before quickly decaying within a few seconds. By contrast, conscious stimuli would be distinguished by their lack of "encapsulation" in specialized processes and their flexible circulation to various processes of verbal report, evaluation, memory, planning, and intentional action, many seconds after their disappearance. [We] postulate that this global availability of information is what we subjectively experience as a conscious state. (Dehaene and Changeux 2011: 210)

This idea develops Baars' (1989) earlier global workspace theory of consciousness according to which content is conscious once it enters the global workspace because there it gains the capacity to be broadcast to a host of higher-level cognitive consumer systems for executive control, motor control, report, and so on.

The global neuronal workspace theory is attractive because it follows the strategy of trying to pinpoint the mechanism in the brain that endows perceptual content with a set of properties considered central to consciousness, including perceptual unity. It is also attractive because in an extremely impressive series of studies, these researchers are delivering empirical evidence about what governs the entry of perceptual content to the global neuronal workspace. What they find is that

> human neuroimaging methods and electrophysiological recordings during conscious access, under a broad variety of paradigms, consistently reveal a late amplification of relevant sensory activity, long-distance cortico-cortical synchronization at beta and gamma frequencies, and "ignition" of a large-scale prefronto-parietal network. (Dehaene and Changeux 2011: 209)

For example, in a study using intracranial electroencephalogram (which measures electrical activity directly from the surface of the cortex), Gaillard and colleagues compared conscious and nonconscious processing of masked visual stimuli and observed a pattern of neuronal activity that "fits with the global workspace model, which postulates that once a representation is consciously accessed, a broad distributed network, involving in particular prefrontal cortex, ignites and broadcasts its content" (Gaillard, Dehaene et al. 2009: 486).

The crucial question here is what "ignition" is and why some contents ignite and others do not. A good answer to this question should also help us understand why perceptual experience is unified rather than not. As far as I can see, the proponents of the global neuronal workspace theory primarily use "ignition" in a descriptive sense, to capture the rapid spread of activity in prefrontal and parietal cortex that is a signature
p. 213 of conscious perception. It is then noted, as in the quote above, that this is just the kind of pattern of ↳ activation we should expect if something like the workspace theory is correct. Notice though that this descriptive approach does not yet explain what ignition *is* or *why* it should sustain perceptual unity.

Dehaene has offered a more substantial explanation of ignition, in terms of evidence accumulation:

Downloaded from https://academic.oup.com/book/4105/chapter/145830336 by Royal Library Aarhus University user on 19 April 2024

> [C]onscious access would correspond to the crossing of a threshold in evidence accumulation within a distributed global workspace, a set of recurrently connected neurons with long axons that is able to integrate and broadcast back evidence from multiple brain processors. During nonconscious processing, evidence would be accumulated locally within specialized subcircuits, but would fail to reach the threshold needed for global ignition and, therefore, conscious reportability. (Dehaene 2008: 89)

A threshold is set for when one out of two or more competing interpretations of the sensory input has accumulated enough evidence in its favour, and when this happens it enters the global workspace. This is a proposal that is kindred in spirit to the idea of using Bayesian, probabilistic tools to account for perception (though strictly speaking it generates different predictions than the prediction error approach, see Hesselmann, Sadaghiani et al. 2010).

Let us attempt to transpose this into the prediction error minimization framework. Global ignition is achieved when a hypothesis about the world is better able to explain away input than its competitor, and therefore achieves higher posterior probability. What counts as being better able to explain away input is modulated by the expected levels of noise and uncertainty. All up, this is what sets the threshold for ignition. Upon ignition, the perceptual content is made available to consumer systems throughout the brain, and can guide action and decision-making, and be introspectively reported. As such it is close to the way conscious perception was discussed in the previous chapter on attention and its relation to consciousness.

This proposal is attractive. Not only does it fit nicely with the global neuronal workspace theory, the focus on disambiguation and uncertainty-reducing inference repeats themes from our discussion of binding and cognitive penetrability. However, this does not yet seem to quite explain perceptual unity in satisfactory terms. Of course, in so far as global ignition is associated with something like re-entrance (cf. the dynamical core theory), it could be claimed to account for perceptual unity. In his wide-ranging work on the self, Metzinger considers a line of reasoning congenial to this:

> The principle is that the almost continuous feedback-loops from higher to lower areas create an ongoing cycle, a circular nested flow of information, in which what happened a few milliseconds ago is dynamically mapped back to what is coming in right now. In this way, the immediate past continuously creates a context for the present—it filters what can be experienced right now, [...] if
p. 214 we apply this idea to the brain's unified portrait of the world as a whole, then the ↳ dynamic flow of conscious experience appears as the result of a continuous large-scale application of the brain's prior knowledge of the current situation. (Metzinger 2009: 30)

This is appealing but sidelines the probabilistic aspect of the explanation, which is crucial to make sense of ignition in the first place. It seems to me that what is needed is a more substantial link between the unity and the probabilistic explanation of ignition. We need to know *why*, from the point of view of perceptual inference, unity would arise. That would provide a deeper, more satisfactory proposal concerning conscious perception, and perceptual unity specifically. The next section will attempt this explanatory task.

Ignition, active inference, and unity

The idea of ignition is important because global dominance suggests unity, and it is this idea we must put into the context of the prediction error minimization mechanism. The first question to ask is why ignition should occur when the hypothesis that best explains away sensory input is selected?

The picture we have painted of prediction error minimization in previous chapters focuses on two facets of error minimization. On the one hand there is perceptual inference and on the other hand there is active

Downloaded from https://academic.oup.com/book/4105/chapter/145830336 by Royal Library Aarhus University user on 19 April 2024

inference. At various points, I have appealed to the alternation between these two processes. It is important to do both, and to get the balance between them right. I think ignition reflects what happens in the switch from perceptual to active inference, and that conscious perception therefore is tied specifically to active inference. I hope to show that this provides a reasonable candidate for perceptual unity in prediction error terms.

Creatures like us maintain a perceptual hierarchy that models a deeply causally structured world. Having hit on a good hypothesis whose predictions efficiently minimize the prediction error bound on surprise, it is a non-trivial matter to then go on to figure out how to engage active inference. That is, there are many ways that changes to the world or one's position in it could interact with a favoured hypothesis about the world to minimize prediction error. The trajectory of the agent is itself a hidden cause of its own sensory input, which can evolve in a non-linear way as the agent engages the environment. Prediction of the flow of sensory input, conditional on agency, is therefore going to involve many assumptions and thus parameters on many levels of the perceptual hierarchy.

Described like this, going from perceptual to active inference is a situation that calls for something like a global workspace. Here is an everyday illustration of the point: you perceive a café in front of you and
p. 215 experience thirst. ↳ Now prediction error can be minimized by getting something to drink from the café, but it can also be minimized by waiting until you get home and having a glass of water there. A whole range of considerations can make this a difficult decision (for those of us who are slightly neurotic, perhaps): is it worth the money, would you end up getting too much caffeine for the day if you go to the café, what is the chance of genial social interactions in the café, what will happen if you're delayed etc.? If you are with a companion you might need to report some of these deliberations verbally, or you might need to call someone if you anticipate being late. For each of these considerations there is a complicated calculation of fictive actions and counterfactual prediction error cost across multiple spatiotemporal scales. Such wide-ranging generation of scenarios thus matters to active inference. That is, it seems that active inference requires something like ignition of a selected hypothesis into a global space where the predictions can be assessed by subordinate (consumer) systems, conditional on some action. If we had no agency at all—if we had no way of changing the way we sample the world in response to prediction error—then it seems there would be no need for ignition into such a global space. Put simply, we can only do one thing at a time and this thing is prescribed by our singular hypothesis about what we are doing. This necessarily entails predictions that are internally unified throughout the hierarchy—in a global sense.

The proposal is then that anointing one hypothesis about the world as the best prediction error minimizer leads to ignition because the system will be using this hypothesis to inform predictions of the flow of sensory input as it acts on the world. Ignition would be the point at which the system, given a specific context, is satisfied with the posterior probability for a hypothesis (or, in Dehaene's terms, when enough evidence has been accumulated) to warrant a shift to active inference and the production of descending proprioceptive predictions that mediate action.

This gives a rationale for the idea of a threshold—ignition—being central to conscious presentation. A threshold is needed to ensure there is a switch from perceptual to active inference, a point at which a model is held fixed and prediction error minimized given the way the world is assumed to be. It makes sense that ignition goes with global broadcasting because it is a complicated task to figure out how to minimize prediction error given action—a task that requires generation of expected flows of sensory consequences along many dimensions. Of course, how the threshold is set will be highly context-dependent: the planned action and its salience will matter, as will the expected precisions of the prediction error and the confidences in the hypothesis and its competitors.

We might even speculate that the reason this global, multimodal hypothesis is *conscious* is that being conscious serves as a reminder of which hypothesis is currently held fixed for active inference. This would

Downloaded from https://academic.oup.com/book/4105/chapter/145830336 by Royal Library Aarhus University user on 19 April 2024

p. 216 resonate with an early idea from Gregory that consciousness (or qualia) "flags the present" (see also ↳ Metzinger 2004: 127; 188). The point here would be that there is a need to flag the present not only because, as Gregory speculates, hypotheses are essentially timeless, or because the present is potentially dangerous, but rather because the system needs to know which hypothesis is currently prescribing action.

With this I have suggested a way to tie together prediction error minimization with the ignition characteristic of conscious experience. They can be related under considerations of active inference. Now to the question that began it all: why is there perceptual unity? Specifically, why would fixing on a model and beginning to generate policies for active inference invariably go with unity?

The unity thesis is that perceptual experience is never split up between distinct phenomenal fields: there is always just one overall phenomenal field that subsumes any conscious perceptual state you have. Now 'phenomenal field' is not an easy notion to make precise and explicit, even though the intuitive idea extracted from introspection is appealing enough. Nevertheless, as I will suggest next, there are reasons to expect that active inference comes with some kind of unity.

The first point to make in this regard is that active inference ties together represented states in a causal manner. This is because active inference relies on the sensory consequences of *intervention* on the states of affairs in the world, and intervention is crucial to extract causal rather than mere associative information. This is one of the central tenets in contemporary accounts of causation (see Pearl 2000: Ch. 1). If there were no active inference at all, then our hypotheses would be much more loosely tied together; they would be conjectured but never verified. As such they would not need to present the states of affairs of the world as being related in anything but a statistical sense (at best causal inference would be serendipitous). This brings us some way towards expecting unity since a deeper causal structure generalizes across events and will in that sense do more to unify. It seems more likely that when we introspect perception that is based on a hypothesis with a deep causal structure it will appear more unified, than when we introspect perception based on a merely associative hypothesis with no causal information.

However, this notion of unity by causal generalization does not yet explain the key notion that there is *just one* overarching causally structured hypothesis. To explain this, we need to appeal to our possibilities of minimizing prediction error in active inference. The idea here is simple: you cannot simultaneously use two or more different, competing causal hypotheses as a basis for sampling the world to minimize prediction error. In active inference you have to rely on just one hypothesis. *Two* people can rely on two competing hypotheses because they can selectively sample each their own hypothesis of the world in the different ways mandated by their respective hypotheses. But *one* person is tied to modelling intervention in the world in
p. 217 terms of one bodily trajectory giving rise to one flow of input. This restriction tallies well ↳ with the notion of the Bayesian brain we rehearsed in Chapter 1–2, you can only select one hypothesis among competing hypotheses—you cannot accept both the null and alternative hypotheses at the same time, on the basis of the same evidence.

This can be illustrated by appeal to binocular rivalry scenarios. If we have *two* individuals and present one with an image of a face and the other with an image of a house, then there is, in a trivial way, disunified perception. One person experiences a face and the other a house. Even though this is a trivial sounding case, we can explain the "disunity" like this: one person can selectively sample the world (via eye movement) on the basis of the hypothesis that it is a face, the other can sample on the basis of the hypothesis that it is a house. Now present the images, one to each of two eyes of the *same* individual, as done in binocular rivalry. Here there is one unified phenomenal field in which the face and the house perception neatly alternate. The proposed explanation for this is that this single individual cannot selectively sample on the basis of the hypothesis that it is a face and, *simultaneously*, sample on the basis of the hypothesis that it is a house. There is only one set of eyes to move around in active inference and so only one causal hypothesis (either "sensory

Downloaded from https://academic.oup.com/book/4105/chapter/145830336 by Royal Library Aarhus University user on 19 April 2024

input is caused by a face" or "sensory input is caused by a house") can be fixed on for active inference at any time.

This then is how we can explain perceptual unity within a prediction error minimization scheme. Perception is *unified* because it is based on hypotheses in a causally structured hierarchical model, and there is perceptual *unity* (that is, one hypothesis only) because the nature of active inference in creatures like us prevents more than one hypothesis from being used as a basis for selective sampling at any time.

If this account of perceptual unity is correct, then we should expect unity to diminish gradually as active inference loosens its grip. Now this is not an easy prediction to test experimentally. I don't think we, in any realistic laboratory situation, can extricate ourselves entirely from active inference so we should probably not expect any introspection of real disunity (as Bayne also insists). Indeed, it may be that as we descend deeply into more inactive inference it is not just unity but consciousness itself that deserts us—if there is no action then there is no reason for ignition. But perhaps there can be conditions under which our confidence in unity should begin to waver as active inference is beginning to be curbed. There seems to be some evidence for this, again from studies of binocular rivalry.

Binocular rivalry can be modulated but not extinguished by endogenous attention to one of the two stimuli. The more one attentively explores one stimulus, the more it tends to dominate in perception. We have construed endogenous attention as a type of active inference so, if decreasing active inference predicts more disunity, we should expect less attention to bring less rivalry. This seems to happen. In a study of binocular
p. 218 rivalry, Zhang and ↳ colleagues directed participants' attention elsewhere and found that alternation seems to stop (Zhang, Jamison et al. 2011; see also Brascamp and Blake 2012). They conclude that attention is necessary for rivalry, which is reasonable enough. The way I prefer to interpret this within the prediction error framework is that as attention is withdrawn there is less active exploration of the stimuli so there is less imperative to subscribe to just one of two competing causal hypotheses and both stimuli are then beginning to be accommodated. Now this is not clearly a case of disunity since it is hard to know whether instead of having a *dis*unified perceptual field the participants in this study just adopt *one* fused hypothesis of the stimuli—that is, a unified hypothesis that ignores the causal structure of the two stimuli. The problem is of course that getting a firm fix on whether there is perceptual disunity would require getting participants in the study to attend to the stimuli, to report them better, and this would reinstate active inference. It is suggestive however that the strong urge towards a unified hypothesis with temporal alternation between causally distinct percepts weakens as active inference (here, in the shape of endogenous attention) is withheld.

One hallmark of the global workspace is that perceptual content in the workspace becomes available for introspective report. Such reporting of conscious experience is a moment in active inference in the sense that engaging in that kind of verbal or button-pressing behaviour is just engaging in prediction error minimization. This means that when an individual is requested to introspectively report on something, then there will automatically be a bias towards perceptual unity. Therefore, if we wanted to discover instances of disunity by curbing active inference, then it would be better to not acquire introspective reports—but of course then it is very hard to know what participants in a study experience. This appears to be a bind.

Another rivalry study is relevant for this apparent bind. Naber and colleagues exposed participants to rivalry between stimuli that moved in opposite directions, asked them to report what they experienced, and meanwhile recorded eye movements (Naber, Frässle et al. 2011). One of their dependent measures was optokinetic nystagmus, which is the rapid back-and-forth movement of the eyes when they are following a moving object (such as when a person looks out a car window). They observed that optokinetic nystagmus follows the reported percepts well. When subjects report seeing a leftward motion then the optokinetic nystagmus would be that characteristic of observed leftward motion. This is of course itself an exemplification of active inference: the hypothesis predicts a leftward motion and the eyes move

Downloaded from https://academic.oup.com/book/4105/chapter/145830336 by Royal Library Aarhus University user on 19 April 2024

accordingly in a selective sampling of the world, constrained by knowledge of head-movement. Next the participants were *not* asked to report their experience with button-presses and instead only the eye movement was recorded during rivalry. The reasoning here was that the eye movement had been shown to
p. 219 be a reliable indicator of the experienced percept, so the rivalry ↳ could be studied without muddying measurement with requirements to introspectively report what was experienced. It turned out that the alternation between optokinetic nystagmus in the two different directions slowed down under these more passive conditions. Hence, when there is less pressure to engage active inference and actively report our experiences there seems to be less need to maintain distinct hypotheses of the world. As before, it is hard to establish unequivocally that there is disunity in this case, but it is suggestive that when report is not required the requirement for unity, where one or the other hypothesis is chosen, seems to drop off.

These studies of binocular rivalry are consistent with the idea that perceptual unity is fuelled by the need for unequivocal selective sampling in active inference. As the requirement for sampling goes down so does the inclination to work with only one model of the world.

Action-based unity and indirectness

Though progress has been made in our understanding of the unity of consciousness (see Bayne 2010), it is still a notion that is difficult to understand and make clear. Much of our grasp of it is sourced in relatively murky introspection. My account attempts to offer fairly concrete mechanisms for what *unites* perceptual experience, namely hierarchical causal structure, and what makes perceptual unity *unitary*, namely the requirements of active inference. To me, this is attractive because it not only describes unity but because it explains why we should expect there to be unity.

When looking at consciousness science more broadly, this proposal fits nicely with the global neuronal workspace theory, proposed by Dehaene and others. It comes with the twist that the contents making it to the global workspace are the ones necessary for active inference. These are representations of the world modelled up against our agency in the world, generating prediction error and thereby initiating action. This means there is a deep connection between conscious perception and unity and action.

This relates to work by Susan Hurley, which does much to connect the unity of consciousness with action. She points out that a dichotomous conception of perception as input to the system and action as the output of the system is not fruitful. A more dynamic concept of perception and action is proposed instead:

> [T]he unity of consciousness has both normative, personal-level and subpersonal aspects. The relations between these levels can be approached via the closely related but more general idea of a perspective: the interdependence of perception and action that perspective involves can be
> p. 220 explained in terms of their ↳ co-dependence on a subpersonal dynamic singularity. This subpersonal aspect of unity does not support sharp causal boundaries either between mind and world or between perception and action. Moreover, it can provide an antidote to the inward retreat of the mind in modern philosophy. At the personal level, the self does not lurk hidden somewhere between perceptual input and behavioral output, but reappears out in the open, embodied and embedded in the world. (Hurley 1998: 3)

Much in this sentiment sits reasonably well with much of what the prediction error minimization account can offer. It seems to me Hurley is right that we cannot understand unity unless we factor in the way action and perception work in tandem. In particular, as I have been emphasizing above too, there would be no need for unity if there were no agency. The prediction error account gives a principled explanation of why something like a singularity (that is to say, I take it, a unity) should arise when perception and action works in tandem, namely in order to minimize prediction error.

Downloaded from https://academic.oup.com/book/4105/chapter/145830336 by Royal Library Aarhus University user on 19 April 2024

The embodied and externalist aspects of Hurley's approach (and by extension that of others in the embodied cognition camp) are much less attractive, from the point of view of prediction error minimization. The prediction of the generative model of the world maintained in the brain is an internal mirror of nature, it recapitulates the causal structure of the world and prediction error is minimized relative to the model's expected states. Similarly, the starting point for the prediction error account of unity is one of indirectness: from inside the skull the brain has to infer the hidden causes of its sensory input, which crucially depends on interaction with the creatures' body. It turns out that such inference requires close tandem work by perception and action, but this is because the inferential task is essentially indirect. As such it is difficult to free it from suspicions of being what Hurley calls a "traditional view of content [that] generates skeptical worries about how knowledge of the world is possible: why couldn't our beliefs about the external world be rampant delusions produced by mad scientists manipulating our brains?" (Hurley 1998: 8).

This suspicion is inevitable for the prediction error account (and was acknowledged and set aside in Chapter 8). Selecting and shaping parameters certainly happen in close causal commerce with the world, but it is nevertheless an inferential process of explaining away the ever changing but oftentimes predictable patterns of sensory input. What are behind the barrier of sensory input are *hidden* causes in the sense that they must be revealed in causal inference. An appeal to action, on the prediction error scheme, reduces to an appeal to inferences about different kinds of patterns of sensory input. If a mad scientist was a hidden common cause of all that sensory input we would have no way of knowing unless she made an independent causal contribution to sensory input. The mad scientist would have a difficult job at keeping up with our
p. 221 selective sampling of the world in active inference but in principle it is ↳ just about projecting sensory data that are close enough to what we expect on our sensory "screens".

It seems to me that it is better to accept the indirectness implications of the prediction error minimization framework than trying to force it into an embodied and externalist shape. It is better, that is, to live with the risk of radical scepticism than to belie the core idea that perception and action are both inferences on the statistical properties of the sensory input.

The final picture is then that our grasp of the world—the way we mirror its causal structure—is at the mercy of the inferential tools we have internally in the brain. This lends a certain kind of fragility to its representations and to the self, which will be the focus of the next chapter.

Summary: unity and causal seclusion

We were after an explanation of perceptual unity from the perspective of prediction error minimization. The explanation begins by noting that conscious perception belongs in the global neuronal workspace, which houses the perceptual content (or hypothesis under an internal model) deemed good enough for active inference. Active inference requires prioritizing just one hypothesis of the world, and this delivers unity in the sense of ensuring there is never more than one subsuming perceptual field. The sense of unity of the conscious field itself was explained in terms of the causal rather than merely associative nature of internal models.

Even though this account relies on the individual's actions in the world to account for facts about perceptual structure, it in fact cements the indirect, inferentialist view of mind and action. This is a theme I will explore in the next chapter.

Downloaded from https://academic.oup.com/book/4105/chapter/145830336 by Royal Library Aarhus University user on 19 April 2024

Notes

Page 208. ["This conception of conscious perception…"] I propose a way to address the problem of consciousness without stumbling into the hard problem. The view I propose essentially is a kind of empirically enhanced commonsense functionalism; it adds empirical information to Jackson's list of core properties, discussed in Chapter 8.

Page 211. ["It would count strongly in…"] The development of the dynamic core theory of consciousness by Tononi in terms of an information theoretical quantity called information integration is also attractive and not entirely alien to probabilistic accounts of perception (Tononi 2005).

p. 222 *Page 211*. ["I want to make one further…"] This global neuronal workspace is described in terms of neural activity and connectivity: "Our proposal is that a subset of cortical pyramidal cells with long-range excitatory axons, particularly dense in prefrontal, cingulate, and parietal regions, together with the relevant thalamocortical loops, form a horizontal 'neuronal workspace' interconnecting the multiple specialized, automatic, and nonconscious processors […]. A conscious content is assumed to be encoded by the sustained activity of a fraction of GNW neurons, the rest being inhibited. Through their numerous reciprocal connections, GNW neurons amplify and maintain a specific neural representation. The long-distance axons of GNW neurons then broadcast it to many other processors brain-wide" (Dehaene and Changeux 2011: 210).

Page 214. ["The picture we have painted…"] The claim is that active inference requires something like ignition because active inference requires promoting one hypothesis as the best prediction error minimizer, and as the hypothesis on the basis of which input in the longer run is predicted. This view seems to me consistent with very recent work by Friston, Breakspear and Deco (2012), who explains the perceptual switching associated with ignition in terms of how the system visits states with good prediction error minimization but then (because such states by definition have flat prediction error landscapes) vitiates those states, which enables (re-)ignition and visiting of new states. This relates to the imperative to shift between perceptual and active inference I have discussed in Chapter 4 and 7.

Page 219. ["These studies of binocular rivalry…"] The prediction error framework is flexible, and unity may be explained in other ways within it. Wanja Wiese and Thomas Metzinger (Wiese and Metzinger 2013) propose that unification of perceptual content happens because not only must individual perceptual processes predict what happens in the world, they must also predict what other perceptual processes know. If all of them are doing this, then part of what they will be representing is how they are themselves being represented. This creates a nested series of representations. This aims at explaining a phenomenological account of unity of consciousness in terms of global self-embedding (or self-representation). The question is of course why such mutual representation should be required on a predictive coding story. Wiese and Metzinger propose that there is such nested mutual emulation because it facilitates the quality, or reliability, of individual representations, and because it could facilitate examinations of how well individual representations fit together, of how coherent they are. I think care must be taken in how these ideas are fleshed out. First, the question of reliability is just the question of precision optimization, and we have seen that that is a matter of second order statistical inference, which does not require knowledge of what the means represent. So ensuring reliability, in this sense at least, does not mandate nested representation of the content itself. Second, the coherence of individual representations is related to the question of sensory integration, and we have seen how this is a matter of precision optimization under top-down causal expectations; this is a story that does not obviously lend itself to the notion of nested emulations. However, perhaps these ideas can be worked out in different ways, to avoid these initial objections. There is something very appealing about the suggestion because it implies that the brain benefits from common knowledge: not only
p. 223 must the ↳ individual representations have mutual knowledge of the same things in the environment, they must also have *common knowledge*—that is know of each other what they know about what they each know, and so on. There are well-described epistemic, agential, and practical reasoning benefits to such common knowledge between independent individuals (Chwe 1999; Chwe 2000). It is therefore not inconceivable, and a rather beautiful idea, that similar processes could be at play within the human brain as are in play among human brains. I briefly explore this theme in the final chapter of the book.

Downloaded from https://academic.oup.com/book/4105/chapter/145830336 by Royal Library Aarhus University user on 19 April 2024