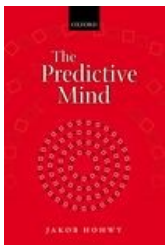

The Predictive Mind

Jakob Hohwy

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

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

CHAPTER

4 Action and expected experience

Jakob Hohwy

https://doi.org/10.1093/acprof:oso/9780199682737.003.0005 Pages 75–97

Published: November 2013

Abstract

The description of the perceptual mechanism has so far been very passive but in fact action is heavily involved in unconscious perceptual inference. This chapter first connects perceptual inference with action via the notion of active inference. An important part of this story concerns the idea that our model of the world includes representation of ourselves. Next, however, some issues arise, concerning the notion of self-fulfilling prophecies, and how surprise is bound under a model of the world. which prompt a more involved and challenging information theoretical approach. This second part of the chapter serves to show why action is so central to our fundamental understanding of why and how we engage in prediction error minimization. This is followed by an exploration of matters arising from this prediction error minimization take on action, including the relation between belief and desire. The chapter ends on a more general note, by summarizing the prediction error minimization mechanism and commenting on why the framework presented throughout Part I of the book is attractive as well as on what challenges it confronts.

Keywords: active inference, information theory, self-fulfilling prophecies, desire, self-organization, homeostasis

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

Collection: Oxford Scholarship Online

A picture has emerged of prediction error minimization as the mechanism the brain uses in perceptual inference. This is in many ways an extremely attractive framework, which includes the following features:

(1) It can begin to deal with the problem of perception. The prediction error minimization framework can respond reasonably to the challenge of providing additional constraints on perceptual inference in a non-circular way, without vicious regress, and without being too intellectualist.

(2) By appealing to the notion of the perceptual hierarchy, prediction error minimization can begin to accommodate aspects of the phenomenology of perceptual experience, such as the mixture of variant

and invariant representation in our first-person perspective.

(3) When the notion of precision expectations, and the corresponding notion of gain on prediction error, is built into the mechanism, it becomes possible to see how different overall processing patterns could arise, and how there can be modulation in the engagement of prior beliefs in different situations.

(4) The framework is extremely parsimonious, with a simple mechanism at its heart, replicated throughout the hierarchy and yet able to fulfil a number of computational functions.

The key idea in all this is to give the brain, skull-bound as it is, access to not only the incoming sensory data but to a comparison between this data and expectations about what the data should be, under a model of the world. The difference between these two is the prediction error, which is then a measurable quantity for the brain, and something that can act as a feedback signal on the way its models of the world are chosen and their parameters revised.

At the end of the last chapter, we noted that this presents a rather worrisome, passive picture of us as perceiving creatures. The picture seems to leave us completely hostage to our respective starting points. It is a trivial observation that we use our powers of agency to improve our position in the world and that we, of
p. 76 course, use the way we perceive the world to inform and guide ↳ agency. Without agency we would be stuck at our starting point, wouldn't be able to improve our situation in the world, and it would be difficult to see why we would be expending energy representing the world in the first place.

This suggests that the prediction error minimization idea should, at the very least, be consistent with the presence of agency. But more than that, perceptual inference should be seen as providing an essential part of what makes us beings with agency. We are engaged in perceptual inference at least in part *because* we need to act on the world.

In fact, however, there is a much deeper connection between perception and agency, which springs from the very idea of prediction error minimization. Perceiving and acting are but two different ways of doing the same thing.

The plan for this chapter is first to connect perceptual inference with action. This serves to strengthen the idea that perceptual inference is much like scientific hypothesis testing. An important part of this story concerns the idea that our model of the world includes representation of ourselves. Next, however, some issues arise, which prompt a more involved and challenging information theoretical approach. This second part of the chapter serves to show why action is so central to our fundamental understanding of why and how we engage in prediction error minimization. This is followed by an exploration of matters arising from this prediction error minimization take on action. This serves to show how perception and action are unified in the same prediction error minimization framework.

The chapter ends on a more general note, by summarizing the prediction error minimization mechanism and commenting on why the framework presented throughout Part I of the book is attractive as well as on what challenges it confronts.

Active inference in perception

Perceptual inference has been presented as a matter of selecting and adjusting hypotheses about the world in response to the prediction error they engender. It follows trivially that the upshot of the brain's prediction error minimization activities is to increase the mutual information between the mind and the world—to make the states of the brain as predictive as possible of the sensory input caused by events in the world. This account has largely suppressed a very obvious point, namely that the mutual information can also be increased by making the sensory input from the world more predictive of the states of the brain's model, that is by changing the input to fit the model rather than changing the model to fit the input.

The notion of prediction error minimization encompasses both directions of fit. That is, the model
p. 77 predictions will also be less erroneous the more the ↳ sensory input is made to fit the predictions. Given the basic idea that the main aim of the brain is to minimize prediction error, we should expect it to exploit this different direction of fit too. That is, we should expect that the brain minimizes prediction error by changing its position in the world and by changing the states of the world, both of which will change its sensory input. This can be captured in the expectation that the brain uses *action* to minimize prediction error.

Indeed, it falls natural to look for a role for action in perception. We are clearly very active in the way we go about perceiving the world. We explore, check out, test, look closer, feel, and so on, all of which are ways of actively engaging with the world and thereby changing the sensory input we receive. This is how I found it natural to describe perception from the very beginning. In Chapter 1, I used as the initial example perception of a bicycle where the posterior probability that it is a bicycle in front of me came about by predicting how the input would change if I *walked around* it and it really was a bicycle rather than a cardboard poster of a bicycle. Finding an example of perception that did not have an active element required the somewhat contrived cases of an individual locked in a room and trying to infer the source of a sound, or the idea of plugging a leaky dam. This was useful because I began by focusing on perception in fairly representational terms. But now action needs to be restored to its central place in our understanding of perceptual inference and prediction error minimization.

Another reason to focus on action in perception comes from the long-standing analogy, stressed as we saw earlier by both Helmholtz and Gregory, between perception and scientific hypothesis testing. Scientific hypothesis testing is paradigmatically a matter of experimentation, that is, active intervention by the scientist in causal chains in order to reveal causal relations. If perception is like hypothesis testing we should expect a similar notion of intervention in perception.

More broadly, this connects to contemporary debates about the difference between merely associationist, statistical inference, and properly causal inference, with the latter being based on interventions where independent variables are manipulated in a controlled fashion. Though statistical associations are necessary for causal inference, there is a limit to how much causal knowledge associations alone can provide us with. Passive observation can allow us to guess at causal relations between random variables *A* and *B*, but it is not until we actively test them that we will know whether *A* causes *B*, or *B* causes *A* or whether they perhaps have a common cause *C* (Pearl 1988; Pearl 2000; Woodward 2003).

So there must be a role for action in the story told so far: mutual information can be enhanced through action and action is obviously involved in perception. The basic idea would be that the brain uses a specific hypothesis about the world, predicts what the sensory input would be like were this hypothesis true, and
p. 78 then actively samples the world selectively to get this ↳ predicted sensory input. In other words, the brain generates a fantasy, a set of predictions that now do not fit with the current sensory input. This induces a prediction error that can be minimized by turning the fantasy into reality, that is, by acting to bring oneself into the predicted situation.

There is an immediate challenge here. Relying on action is relying on making the sensory input fit with our expectations, which seems to turn the proposal into a more unattractive analogy to *bad* scientific hypothesis testing. If the brain can do its job by using action to make the world fit its expectations, then it should just adopt expectations that are easy to make come true. For example, if you predict darkness then you will have minimized prediction error very nicely by closing your eyes. Clearly, in the long run, adopting such a strategy will not be beneficial. If the tiger is approaching and you minimize prediction error by closing your eyes, your success will be short-lived—and you may experience other prediction errors, when your predictions about not being eaten are violated. We therefore need to figure out how to accommodate this direction of fit for prediction error minimization without making the whole account implausible.

Luckily, it is easy to answer this challenge. Recall that simple examples of Bayesian inference rely on ordering hypotheses or models according to their prior probability, and then weighing this with the likelihood of a given model actually producing the sensory input in question. Here prior probabilities provided the needed additional constraint on that kind of inference. In the present case, additional constraints are also needed lest we adopt poor models that are too easy to confirm by changing our sensory input in the myopic and implausible ways exemplified above. The difference is that this time, the system needs to use the hierarchical hypothesis with the highest posterior probability and project it to actively test the world according to it.

The process is therefore first to rank hypotheses according to their posterior probabilities, delivered from perceptual inference of the type described in the previous chapters. Then the system actively samples the world to see if new sensory evidence can be produced, which is in accordance with the preferred hypothesis. In other words, we should selectively sample sensory input that conforms to our predictions; where, crucially, we predict that these sensations will minimize the uncertainty that persists about our hypothesis. If predictions are confirmed then the cycle continues—otherwise, the hypothesis generating those predictions is discarded in favour of a more plausible hypothesis that better minimizes prediction error.

For example, the brain receives some sensory information and by using Bayes' rule, implemented with prediction error minimization, it ranks the hypothesis that it is looking at a face higher than any other hypothesis, given the input. Clearly, it will have failed in prediction error minimization if it sidelines this hypothesis in favour of another, for example, that it is observing utter darkness. This is why active testing of
p. 79 hypotheses should proceed on the ↳ basis of predictions from the hypothesis with the currently highest posterior probability, in this case the face hypothesis. This rules out the implausible and dangerous case of minimizing prediction error in action by adopting the hypothesis that it is dark, then predicting darkness and successfully closing one's eyes to confirm this. That hypothesis has a very low posterior probability (unless it is bedtime) and will therefore not be a good basis of active sampling.

The situation is then this. Perceptual inference allows the system to minimize prediction error and thus favour one hypothesis. On the basis of this hypothesis the system can predict how the sensory input would change, were the hypothesis correct. That is, it can test the veracity of the hypothesis by testing through agency whether the input really changes in the predicted ways. The way to do this is to stop updating the hypothesis for a while, and instead wait for action to make the input to fit the hypothesis. If this fails to happen, then the system must reconsider and eventually adopt a different or revised hypothesis. For example, if the highest posterior goes to the hypothesis that this is a man's face seen in profile, then the system may predict that by moving visual fixation down towards the chin, a sample will be acquired that fits with this hypothesis. If it does, then this further enhances the probability that this is a man's face; if it does not fit then the system may have to go back and revise the hypothesis such that it expects the cause of its input to be, say, a woman's or a child's face (for computational modelling of face perception, see Friston 2012; Friston, Adams et al. 2012).

The question arises, why does the system need to engage in this kind of active inference if it has already settled on a hypothesis as having the highest posterior probability? What more is there to do than ranking hypotheses? There are two answers to this, which both have to do with reducing uncertainty, that is, with prediction error minimization.

The first answer is that in these cases, action enhances the posterior confidence in the inference. Action makes decent inferences better. For example, I am more confident I am looking at a man's face after successful active sampling of the world according to this hypothesis. This helps decrease uncertainty especially in cases where the winning hypothesis did not have a very much higher posterior than its competitors at the outset. In other words, action can help create much stronger minimization of prediction error than mere passive observation.

In this sense, action can make predictions more reliable—it can make the hypothesis stand out more distinctly against its competitors. Mere perception alone cannot do this, it is hostage to the whims of the incoming sensory data and cannot focus on one hypothesis and ask whether it is really true. As usual, Helmholtz aptly anticipates this idea: "We are not leaving ourselves passively only to the [sensory] impressions intruding upon us, rather we *observe* ["*beobachten*"], that is, we bring our organs into those
p. 80 conditions ↳ under which the impressions can be most precisely distinguished" (Helmholtz 1867: 438).

Notice that there is nothing un-Bayesian about this. Bayes' rule tells us how to update belief in the light of new evidence. What we do here is update belief in the hypothesis, which previously got the highest posterior, in the light of new evidence, namely the evidence attained by, for example, actively moving our eyes around.

The second answer to the question why the system should bother with processing of already favoured hypotheses is that it is efficient and quick to do so. What I have described so far sounds slow and laborious: I sit passively for a long time, amass as much sensory evidence as possible, patiently rank my hypotheses according to their priors and likelihoods, and then I actively test the best one by sampling the world according to its predictions. But often it is much quicker to form a quick impression of what the ranking might be and then actively test the hypothesis I merely surmise is best. In active testing I can pick a prediction that is made very likely by the hypothesis but very unlikely to occur by chance. If this prediction holds, then the likelihood term is weighted highly and the posterior probability is reinforced. In contrast, in passive observation I have to wait for observations to occur that are strongly predicted by the hypothesis.

There is a related, more systematic reason why action can help reduce uncertainty. In many cases of contextual interaction and other causal relations, observation alone will not distinguish between hypotheses where there is causation between two random variables and hypotheses where there is a common cause of covariation amongst the variables. The statistical associations can support either hypothesis equally well and only differences in prior probability allow one to get a higher posterior. Even though there is a favoured hypothesis in these conditions, it may not be favoured strongly. This leaves the system lacking in confidence that its inference is reliable. This is a type of uncertainty that can be efficiently reduced by intervening actively. For example, I can rule out that there is a causal relation from *A* to *B* if intervening on *A* fails to invariantly change *B*. Given that we can in fact engage in causal inference, and given that observation alone cannot distinguish between such causal models, we can see that we must be relying on intervention, that is, on active inference.

What we have so far is this. If the system can act on the world to change its own sensory input, then it can test its own hypotheses. It can do this in a Bayesian way by testing primarily those hypotheses that have high posterior probability, endowed from passive perceptual inference. Conditional on the evidence attained in action (for example, as one's eyes move around) a given hypothesis can increase its posterior probability. Through action, already selected hypotheses can be made much more reliable in the sense that they

p. 81 minimize prediction error very efficiently. Action in this sense of testing ↳ perceptual models is therefore a moment of prediction error minimization—it is *active inference*.

Modelling the agent, and acting

Putting action in these Bayesian terms makes it sounds as if there is a clear-cut distinction between perceptual and active inference. This is true in the sense that they are associated with very different functional roles: they have different directions of fit between models and sensory input. It is also very clear that the system must keep distinct its updating of models from its acting on those models to sample the world. But it is not as if the system needs to have distinct periods of inactivity and activity. As long as the functional roles associated with direction of fit are respected, we can accept that the system is moving around most of the time. Thus, viewed as a whole, the probability of getting a certain sensory input is conditioned on causes in the world jointly with the actions of the system itself. This allows perceptual inference to take into account change in sensory input that is due to the creature's own action.

This point is important because if the system has agency, then it can interact as a hidden cause with other hidden causes in the world. Therefore the system's model of the world needs to include a model of itself and its trajectory throughout the world just as it needs to model other hidden causes.

In this sense there is not much difference between the system's model of itself and its model of other interacting causes, such as the cat and the occluding picket fence we considered earlier. Just as there is a non-linear relation between cats and fences, there is a non-linear relation between acting systems like us and objects in the environment. For example, the simple occlusion of the cat by the fence can be modulated in many different ways by the perspective of the perceiving system (for example, the inference that it is a cat behind the fence takes into account the agent's shifting perspective as he or she walks past).

Here it is again necessary to invoke the perceptual hierarchy. Agency happens at many different time scales, from the very short micro-saccades where the eyes quickly fixate a new place, over hand-reaching movement, to long-term endeavours like climbing a mountain. Actions at these time scales each have different, interacting effects on the sensory input the system will receive. Therefore the internal representation of the agent needs to accommodate this hierarchy such that the effects of its actions on its sensory input can be predicted; in other words, so that its own movement does not inadvertently increase prediction error.

It is hard to resist the temptation here to think of this multilayered, internal model as in some sense a model
p. 82 of the *self*—of who and what the agent is ↳ (for discussion, see Metzinger 2009). I will indeed briefly pursue this intriguing thought later, in Chapter 12. The important point for now is the requirement that the internal, generative model needs to include a model of itself such that it can explain away its own sensory input, even when changes in that input are partially caused by itself. This means that the system can learn, and come to expect, patterns in how its sensory input would change given certain actions. For example, it can come to expect how things would change if the eyes fixate in a certain way, given it is a face one is looking at. These learned patterns can fuel action because they are essentially predictions of what the sensory input will be, given a generative model that includes parameters for hidden causes in the world including the hidden cause that is oneself.

Now consider how action comes about in a system that models itself and which can act. The representations of predicted sensory input are counterfactual in the sense that they say how sensory input *would* change if the system *were* to act in a certain way. Given that things are not actually that way, a prediction error is induced, which can be minimized by acting in the prescribed way. The mechanism for being a system that acts is thus nothing more than the generation of prediction error and the ability to change the body's

configuration such that the antecedent of the counterfactual actually obtains and error is suppressed. Action therefore does not come about through some complex computation of motor commands that control the muscles of the body. In simple terms, what happens is instead that the muscles are told to move as long as there is prediction error. The muscles of the body are thus at the mercy of the prediction error generated by the brain's model of the way the world is expected to be like but isn't. Prediction error is then the simple mechanism that controls action.

An immediate objection to this story about what generates action is that it implies great variability in our routes to distal goal states. You might predict what your sensory input is like if you have moved your arm from location *A* to location *B* (or engaged in more complex action such as climbing a mountain) but there will be multiple ways of getting to the goal-state. In fact, there is in principle an infinite number of ways the organs of the body can be brought into any given condition. And yet we are able to move swiftly and in fairly uniform ways. The response to this worry must be that the brain represents not only the goal state but whole flows of expected sensory states, that is, how our sensory states will change as behaviour unfolds. In this way, behaviour can be controlled by prediction error in an online fashion, leaving little gap between the present state and the goal state.

Before moving on, I will make four brief comments to support this idea of action controlled by expectations for the flow of sensory states.

First, the modelled flow of sensory input concerns not only the kinds of exteroceptive input we have been focusing on so far, namely visual, auditory, and tactile input. It also concerns interoceptive input such as
p. 83 states of arousal, ↳ heartbeat, and proprioception and kinesthesis. This means, for example, that if proprioception is not as predicted, then the body is not configured in the right, predicted kind of way, and prediction error will then persist until reflex arcs have successfully fulfilled the predictions. Hence prediction error can control the body directly, in virtue of these more inner sensory channels. I will discuss this sort of control in abstract terms but notice that it essentially concerns reflexes and homoeostasis that are fundamental for survival, and that can be cast easily in terms of minimizing prediction errors or deviations from set points. A key future, explanatory task for the notion of active inference lies in connecting the ideas of sensorimotor predictions to more long term, abstract notions of motivation and action.

Second, at very low levels in the control hierarchy, which oversees movement at short time scales, it is likely there is a very restricted and therefore automated repertoire of parameters for how the body could be configured. This will facilitate inference because the expectations for how the sensory input changes will be harnessed in such simple reflexive patterns rather than having a full range of possible movement patterns to choose from. This may not be so different from the way perceptual inference at low levels seems to rely on rather restricted classes of model parameters (for example, what seems to be expectations for a restricted range of line orientations for different cells in early visual cortex).

Third, even though action is controlled through prediction error minimization based on expected flows of sensory states there may also be room for exploratory behaviour. Thus it is possible that movement sometimes begins with apparently random "jittering" or itinerant wandering about in different directions to figure out which direction produces best minimization of prediction error. This direction will then be favoured and will eventually lead to the goal state.

Fourth, this overall account creates a puzzle about how action is triggered, that is how the agent shifts from perceptual to active inference. This is because there will be competition between assessment of the actual proprioceptive input and the counterfactual proprioceptive input. Rather than changing the world to fit with the counterfactual predicted input, the system could just adjust its proprioceptive prediction in the light of the actual input—it could realize that it is not actually in that state. This would prevent action from arising.

A mechanism is thus needed to ensure agency. One intriguing proposal is that this mechanism is attentional, focused on the precisions of proprioceptive input(Brown, Adams et al. in press). Briefly put, action ensues if the counterfactual proprioceptive input is expected to be more precise than actual proprioceptive input, that is, if the precision weighted gain is turned down on the actual input. This attenuates the current state and throws the system into active inference. Being an agent then reduces to a matter of optimizing expected precisions of proprioception, which is a far cry from our commonsense idea
p. 84 of ↳ what makes an agent. If this is correct, then active vs. passive movement should be marked by attenuation of self-generated sensory input. There is evidence for this in many domains, such as our famous inability to tickle ourselves (Blakemore, Wolpert et al. 1998). This tickle effect ought then to be a very fundamental aspect of being an agent, and not something that can be easily upset by more superficial changes in how the tickling sensation is experienced. Recently, George van Doorn, Mark Symmons, and I (ms) found evidence of this by observing how the tickle effect can survive very extensive, unusual changes in body image: you cannot tickle yourself even if you have swapped bodies with someone else.

Bounding surprise

What we have thus far is a fairly simple account of how prediction error minimization can explain action when the direction of fit is that sensory input is changed to match predictions. Action enhances the reliability of favoured models of the world and ensues when prediction error is minimized for expectations of the evolution of sensory input, under a generative model of hidden causes that includes the acting system itself.

There is something slightly odd about this proposal. It is phrased in terms of testing predictions through sensory sampling of the world but the process seems more like engaging in self-fulfilling prophecies. For example, the system's prophecy is that it is viewing a face, this prophecy induces a prediction error that causes the system to selectively sample the world until the error is minimized. That is, by the very act of prophesying that it is a face, the system will do what it takes to bring itself into a condition where the prophecy is fulfilled.

Building perception on a basis of self-fulfilling prophecies sounds wrong-headed. However, at the level of proprioception and interoception, it is the very stuff of survival—it is the basis of physiological homeostasis and biological self-organization (e.g., maintaining body temperature and heartbeat); I will discuss this more later in this section. At the level of exteroception, we have already seen that even if there is such an element of self-fulfilling prophesying, this is not immediately damaging. Firstly, the prophecies on the basis of which the world is sampled are not pure, unfounded prophecies, or wishful thinking. They are hypotheses about the world with evidence in their favour. Secondly, the prophecies are not guaranteed to be self-fulfilling. It may be that the world does not cooperate to satisfy the predictions. For example, as I visually sample what I think is a face emerging from the bushes, I might encounter surprise—where the eyes should be there are just leaves. In that case I could persist *ad nauseam* until the prediction error is
p. 85 eradicated, for example by asking a friend to stick her head out ↳ from among the bushes in the right location. But the brain more often revisits perceptual inference, readjusts in the light of the new sensory input it has generated in active sampling, and elevates a new hypothesis, which then becomes the new and better prophecy.

However, there is a deeper level of understanding, on which it is much less obvious that the worry about self-fulfilling prophecies can be met in this straightforward way. To see this, recall from Chapter 2 that the prediction error minimization framework has been presented in terms of a computational mechanism that works with the aim to minimize surprise (that is, the long term, average surprisal or negative log likelihood of the sensory input). Because surprisal cannot be assessed directly, the mechanism manages to do its job by

generating predictions and minimizing prediction error. Mechanistically, this is done by suppressing prediction error at multiple levels of the temporally ordered hierarchy.

This suggests that the mechanism should be able to change the surprisal. However we noted earlier that, as long as we operate with strictly passive, perceptual inference it is in fact unclear how surprisal itself can be made to change. Surprisal is a measure of how surprising it would be to observe the system in question being in certain conditions, or having a certain sensory input. It is clear that this can only be assessed relative to its normal state, the state we are most likely to find it in. This quantity cannot be changed by perceptual inference because perceptual inference changes the hypotheses about the sensory input and not the sensory input itself. Crudely put, perceptual inference can make you perceive that you are hurtling towards the bottom of the sea with something heavy around your feet but cannot do anything to change that disturbing sensory input which is fast taking you outside your expected states. In this sense, ironically, perceptual inference on its own is impotent as regards what was stated as its main purpose.

The obvious candidate for changing surprisal is agency. A creature endowed with agency can make useful predictions and act on its environment and its position in the environment to ensure it stays within the expected bounds. A more unlucky creature who cannot act would in principle be able to represent its environment but could not change the input it receives.

But is agency, in the sense we have described it so far, really sufficient to minimize surprisal? Action in the form of selective sampling on the basis of probabilistically favoured models can *change* the surprisal by changing the sensory input, but it is difficult to see how selective sampling can minimize surprisal. The problem is that surprisal is defined in terms of the expected states of the creature and if the creature is found outside of those states, then selective sampling cannot bring it back, it can only use active sampling to make its model of the high surprisal environment more reliable.

There seems to be only one way to solve this problem. The creature needs to be endowed with prior beliefs
p. 86 that tie it to its expected states. If it chronically ↳ expects to be in what are in effect its low surprisal states, then it will sample the world to minimize prediction error between those expectations and the state it actually finds itself in. To the extent it is able to minimize this error it will minimize surprisal, though of course it may be so far from low surprisal that it cannot make its way back (for example, deep in the sea with something heavy around its feet). These expectations are defining of the creature, because they tell us its expected states and thereby its phenotype.

There is no doubt that this idea is an ambitious and challenging part of the prediction error minimization framework. It asserts that at some level of description all creatures of the same phenotype share the same prior beliefs about what their sensory input should be and that this explains why we tend to find these creatures in certain states and not others. But, in a circular sounding way, the idea also asserts that the fact that we tend to find these creatures in certain states and not others explain why they have the expectations they have. Intriguingly, the upshot is that phenotypes are predictors (models) of their low surprisal states. Creatures chronically expect to be in those states so they must have a model of them on the basis of which they can generate predictions of sensory input that will maintain them in low surprisal. This is the tenet of self-organization and the "good regulator" hypothesis proposed by Ashby and colleagues nearly half a century ago, namely that from a formal point view, a system like the brain that maintains minimal entropy (or surprise) of an environment must model its environment (Conant and Ashby 1970).

It is again tempting to describe this in terms of self-fulfilling prophecies. Because of who we are we expect to be in certain states. So we prophesy that we will be in those states, and by the very act of prophesying those states we induce a prediction error that causes us to end up in those states. By predicting it we make it so. A related objection is that it presents us as fundamentally conservative creatures. We are inexorably drawn to the unsurprising states we expect to find ourselves in, never to new and exciting states. This may

sound obviously false, since we have explorers, thrill-seekers, and curious people among us. It may also sound obviously false because it seems to predict we will rather forego normal, non-thrilling pleasures of life such as a good meal and a cocktail party in favour of a dark room (Friston, Thornton et al. 2012).

But these objections miss that it is in a rather trivial sense that we are conservative. Consider all the possible states we could be in, where these states can be defined in terms of their causal impact on our senses. There is no doubt it is more probable on average to find us in some of these states and not others (we are rarely going down in the sea with something heavy around the feet). To say that we are fundamentally conservative is then just to say that we tend on average to be found in some sensory states and not others. If we were not conservative in this sense we would be expected to be found in all sorts of conditions, which we clearly are not. A similar response applies to the objection concerning self-fulfilling prophecies. If we could
p. 87 make any old ↳ prophecy and if they were all self-fulfilling, then we would expect to find us distributed across all states of the world.

All sorts of exploratory, thrill-seeking, and curious activities are consistent with this kind of conservatism. It seems likely that we expect that in order to remain in low surprisal states on average we need to engage in exploratory behaviour even if such risky behaviour temporarily increases surprisal. For example, in order to protect myself from wind and weather I might explore different kinds of clothing material and lodgings, not always with luck.

Exploratory behaviour seems especially called for in creatures like us, with deep, complex perceptual hierarchies embodying long time scale representations. In order to obtain a distal goal state, highly complex and context-dependent expectations of sensory flows must be learned. For example, in order to test a scientific hypothesis, or in order to climb a high mountain, a long series of sensory inputs will be expected each of which can be confounded in innumerable ways by other sensory contingencies. It may be that learning such complex priors is facilitated by exploratory, itinerant behaviour.

Similarly, this kind of conservatism does not predict that we will seek out dark rooms over all other activities. If we were dark-room creatures, who were expected on average to be found in dark rooms, then that would be what we would expect and hence gravitate towards (if we were like marsupial moles, this would be our story, perhaps). But *we* are not defined by dark-room phenotypes, so we don't end up in dark rooms. We end up in just the range of situations we are expected to end up in on average. It is true we minimize prediction error and in this sense get rid of surprise. But this happens against the background of models of the world that do not predict high surprisal states, such as the prediction that we chronically inhabit a dark room (I return to the dark room issue in Chapter 8, where I relate it to the philosophical debate about misrepresentation).

This discussion began by worrying about the air of self-fulfilling prophesying in the initial story about the role of action in prediction error minimization. The worry could be dealt with in fairly simple, Bayesian terms but prompted a deeper and much more challenging and ambitious framework in which to understand prediction error minimization.

It may be tempting to keep the simpler, more straightforwardly Bayesian account of perceptual and active inference and leave aside the deeper story about minimizing surprisal through self-fulfilling prophesying. In many respects, the remainder of this book can be read in this less ambitious vein. This is because I primarily use the simpler story about the neuronal prediction error minimization mechanism when I apply the framework to problems in cognitive science and philosophy of mind. However, we cannot *understand* how this mechanism works without the idea that the brain does what it does because it needs to minimize
p. 88 surprisal. Specifically, prediction error bounds surprisal and the only reason for ↳ minimizing the error through perception and action is that this implicitly minimizes surprisal.

This issue goes to the heart of why this framework is so attractive. The prediction error framework builds on the mathematical idea that prediction error bounds surprisal and this is what yields a tractable target for a system like the brain. As expressed in Chapter 2, all the brain has to do is to minimize the divergence between probability distributions (or density functions) given by the sensory input predicted given its generative model and that given by the sensory input (or the recognition model). This sounds sophisticated and complex but can be achieved in a mechanistically simple way by organizing neuronal activity such that it in the most efficient way counteracts sensory input, on average and at multiple levels of the cortical hierarchy. For the first time we can see and describe in precise mathematical terms not only what the brain needs to do, we can also see that this is something the brain can actually do, as the neuronal machine it is. This is what we lose if we throw out the more ambitious story about surprisal, phenotypes, and self-fulfilling prophecies (for computational models and more background, see Friston and Stephan 2007; Friston, Daunizeau et al. 2009; Friston, Daunizeau et al. 2010; Friston 2012).

At this stage it is clear that one could embark on an exploration of some of these ideas in terms of adaptation and fitness. How is it that some creatures end up with the phenotypes they have, why do we have different species, why do some species have more exploratory behaviour and deeper cortical hierarchies than others? (For discussion relating to evolution, see Badcock 2012). It is also obvious there can be discussions of the genetic bases for the kinds of expectations that help keep us in low surprisal states, as well as of what those expectations might be. Similarly, it is tempting to expand into discussion of the mentioned ideas of self-organizing, dynamical systems (for an early statement, see Ashby 1947).

I will not engage these types of question directly here. I am primarily interested in what the account says about our understanding of the world and our place in it as perceivers and agents. For this project, we mainly need the following ideas: the brain is only concerned to minimize prediction error; prediction error can be minimized in perceptual inference, where hypotheses about the world are updated in the light of their ability to predict sensory input; and prediction error can be minimized in active inference, where the confidence in hypotheses is updated in the light of the way sensory input can be brought to fit their predictions. In short, we update our models of the world in the light of sensory input, and sample sensory input in the light of our models of the world. When harnessed in an organ like the brain, with an appropriate hierarchy, with a structure for message passing between levels, and ability to change in response to changing input (i.e., plasticity) this can explain the nature of perception and action in a unified way.

p. 89

Active inference: matters arising

Action is described as an inferential process because, just as perception, it proceeds by prediction error minimization. Hypotheses with high posteriors are strengthened probabilistically if the selective sampling of their consequences for sensory input come out true. Acquiring those samples implies that the sensory system in question moves around or changes the environment, which is action. In this section of this chapter I note some topics of interest that arise from these ideas. These topics anticipate discussion in subsequent chapters.

Desire and belief. Action is mostly described in terms of how we act in accordance with our wants, intentions, and desires, given our beliefs. The question then arises how the description of action in terms of prediction error minimization can accommodate those kinds of mental states. This question is not easy to answer quickly. What drives action is prediction error minimization and the hypothesis that induces the prediction error is a hypothesis about what the agent expects to perceive rather than what the agent wants to do. If this idea is expanded to standard examples of desires, then desiring a muffin is having an expectation of a certain flow of sensory input that centrally involves eating a muffin. This means the concept of desire

becomes very broad: any hypothesis associated with a prediction error can induce a want or an intention or a desire, simply because such prediction error in principle can be quenched by action.

What makes the desire for a muffin a desire and not a belief is just its direction of fit. Both are expectations concerning sensory input, and the "motivator" is the same in both cases, namely the urge to minimize prediction error. For action specifically, it is not obvious that a notion of reward, value, or utility is needed to explain action. The way we learn to act through reinforcement can be explained in terms of prediction error minimization rather than a desire for reward (Friston, Daunizeau et al. 2009). It is then tempting to say that strongly desired states are states associated with much and reliable prediction error minimization. There is an element here of mere redescription. Reinforcement learning and optimal control theory are not substantially revised by the prediction error minimization framework. Rather, their notions of reward and cost-functions are shown to be absorbed into the priors of the prediction error minimization mechanism. The point of this is to unify perception and action and to show how one mechanism, namely prediction error minimization, can account for both.

It may seem odd that all action reduces to a kind of inferential process akin to perception. But this way of putting things is in fact a little disingenuous, since we could just as well have said that perception reduces to a kind of agency. In fact, early formulations of some of these ideas came from computational theories of
p. 90 motor control, which can be generalized to ↳ encompass perception (Kawato, Hayakawa et al. 1993; see further work on forward models, e.g., Wolpert, Ghahramani et al. 2001). This worry is then somewhat misguided because the starting point is that the system in question needs to minimize its prediction error, and then the observation is made that this can be done in different ways. There is a reason for sticking with the sensory idiom, however. The key element in both ways of minimizing prediction error is what happens at the sensory interface between the mind and the world (either exteroceptively or interoceptively).

I think one price to pay for this approach to desire is that in the most fundamental sense we do not get to choose our desires—rather, our desires chose us. Our phenotype determines the kinds of states we expect to be in on average, and these expectations ensure, through the inferential process of prediction error minimization, that we end up in those states. I think this price is worth paying because there is ample scope for individual differences in the many highly context-dependent state trajectories we need to learn and chose among to obtain those fundamental goals. Our individual starting points and our learning histories are different and this predicts many different choices of strategies to get our distal goals. As we saw, this allows both exploration and avoidance of low sensory input states, like the dark room.

The picture of the human mind that goes with this appears very *indirectly* related to the world. As mentioned, there are many different ways to achieve distal goals and so agency depends on learning ways to get there, ways that are consistent with long-term average minimization of prediction error. In active inference, this calls for prior beliefs about the flow of sensory input one can expect to receive. For both belief and desire, what matters is how things look 'from the inside': as long as the sensory input is as expected and internal prediction error is minimized it matters little whether a state is a belief or desire, or what the external world is like, or what things we desire. In this sense organisms like us do not really "appreciate" what the model represents or what things we expect to happen. We do not purposefully aim to represent the world or to desire the world to be in a certain way. We just minimize prediction error and thereby attain beliefs and desires. However, though this is a very 'internal' perspective on the mind, it does not come with the disconcerting notion of self-fulfilling prophecies that drift apart in costly ways from what is true. When we update hypotheses and act on the world we are deeply in tune with the states of affairs in the world—this is what the prediction error construct gives us. Prediction error increases when we organize the brain, our bodies, or the world poorly and that forces us away from those states.

Balancing perceptual inference and active inference. In this chapter a particular vision of action is presented, based on existing neurocomputational theories. The underlying mechanism is one of prediction error

minimization, though with a different direction of fit than we saw for perceptual inference. Now I want to
p. 91 draw attention to the combination of action and perception, ↳ and note what seems a very fundamental combination of processes emerging from the prediction error minimization ideas.

Prediction error is a bound on surprise that organisms can assess and try to minimize through perceptual inference (revising the models and predictions of the world). Perceptual inference can lead to a tight bound on surprise but cannot itself reduce the surprise. Conversely, action can reduce surprise by compelling us to occupy unsurprising states but cannot itself make us select good hypotheses about the world (since the hypotheses stay the same for the duration of action). Acting on beliefs does not ensure the truth of the beliefs but can reduce the uncertainty about them by minimizing prediction error.

The different directions of fit suggest that to optimally engage in prediction error minimization, we need to engage in perceptual inference and active inference in a complementary manner. Perceptual inference is needed to create a tight bound on surprise, so that active inference has a sound starting point. This is obviously crucial since engaging in active inference on the basis of an inaccurate conception of what the world is like is unlikely to land the organism in unsurprising states in the long run. If you want to dodge the bullets rather than stay put you'd better know exactly when the shots are fired.

Action alone is not enough. Often action induces new prediction error, because the world is a complex and uncertain place. Therefore it pays to suspend active inference and revert to perceptual inference so that the bound on surprise can be readjusted, before action is again taken. As you are hit by a surprise bullet you might pause, briefly, to update your model of when they are fired and then reconsider your strategy for action. Perception alone is not enough either. In the long run it is not efficient to aim for only correctness of our generative models—it will do nothing to place us in low-surprisal states.

We should therefore aim to alternate between perceptual and active inference. This alternation of inferential activity seems to me a very fundamental element of who we are and what we do. Getting the weighting of these inferential processes right is crucial to us: if the bound on surprise is not minimized enough by perceptual inference, then action suffers. If we persist with minimizing the bound for too long before we act, then we become inactive and end up spending too much time in states that are too surprising in the long run. If we persist with active inference for too long without pausing to revisit perceptual inference, then inaccuracy mounts and action becomes inefficient. If we react too soon to mounting prediction error during active inference then we get lost in overly complex and detailed models of the world.

It is easy to imagine that people differ significantly in how they manage this delicate alternation between perceptual and active inference—and that this variability depends upon the precision of beliefs about active sampling of the world. Different genetic set-ups, different developmental and learning patterns, and different contexts can sway us towards quick action, slow learning, and different oscillations between them.
p. 92 Perhaps such matters are ↳ involved in some developmental disorders and mental illnesses, such as autism, where patients seem to get stuck in perceptual inference, and schizophrenia where patients may operate with less than optimal generative models. We will revisit some of these questions in Chapter 7.

There will be an intricate relation between how perceptual and active inference is dealt with and other processes that emerge from the prediction error minimization framework. One such process concerns expected precisions. The extent to which we expect precision in prediction errors determines (modulo other contextual factors) how deeply we sample the world and how much we rely on prior beliefs. Hence, if expected precisions are far from optimal, then we may over- or under-sample the world and fail to engage active inference in an efficient manner. Another process concerns optimizing the complexity of models and how to balance complexity against their accuracy of internal models. We touched on this issue in the discussion of overfitting in Chapter 2. A very accurate model of the world will have states and parameters for every little thing that happens. But such a model will be extremely complex, expensive to maintain, and it

may end in overfitting. In other words, it will fail to generalize to new settings and, over time, produce more prediction errors. Moreover, many of the represented states of affairs will be irrelevant for subsequent active inference. So it pays to deploy some version of Occam's razor such that the simplest model that will facilitate active inference in the long run is chosen without oversimplifying so much that the sensory input during action becomes poorly predicted. This requires us to engage in some kind of Bayesian model selection, which has an implicit penalty for complexity. Again, getting the balance of complexity and accuracy right will be a complex, context-dependent affair.

Our current state of mind will depend on how we attend to all these dimensions of perception, action, precision, complexity, and context-sensitivity, and on how the combination of them all enables prediction error minimization in a particular situation. As a very rough characterization of a normally functioning mind, the average mental state would be such that if engaged in action, perceptual re-assessment is never far away, and if engaged in perception, action is often imminent; the fineness of perceptual grain is rarely extreme nor exceedingly abstract; and our perceptual and active inferential processes are mostly drawn towards precise rather than imprecise sensory input.

Prediction error minimization: challenges

With these last comments, the discussion of action is concluded, and together with the previous chapters of
p. 93 this Part I of the book, I have brought everything ↳ to the table we need for the discussions in the remainder of the book. It presents an attractive package, which obviously has great explanatory potential and can be explored in a wide variety of ways.

There are of course challenges to the prediction error minimization framework as presented so far. These challenges cluster around two themes, which I will address in turn below:

1. What is the *evidence* that the brain is (only) engaged in prediction error minimization? How can the idea be tested? What does it predict? Why should we believe it?

2. How *much* can the prediction error minimization principle explain? *How* (given its level of generality) does the prediction error minimization scheme explain anything?

A primary challenge for the framework is accommodating the multitude of different empirical findings from the many different fields of research it applies to. The danger here is that the framework is phrased in such general terms that it is easy to fit in all sorts of evidence in a just-so manner. By assuming convenient priors and likelihoods it is possible to make Bayesian "sense" of pretty much any kind of behaviour ("why do some people commit suicide? Because that's how they minimize surprise"). This means that the theory can appear unfalsifiable.

A similar problem could be mounted for the most abstract formulation of evolutionary theory. But this does not make us think Darwin was fundamentally wrong. There are mountains of evidence in favour of evolution coming from its more specific predictions (and the theory has itself evolved in the light of the data) (Dawkins 2009).

Something similar potentially holds for prediction error minimization and the more general idea of the free energy principle (which basically says that creatures act to minimize the amount of prediction error over time). It will be tested in specific terms, based on quantitative predictions from specific computational models. Of course, such testing has already begun. I have mentioned quite a few studies in the previous chapters and notes, and there are further, concerted efforts to, for example, use dynamic causal modelling in brain imaging to discern whether the brain deals with prediction error in associative learning in the way predicted by prediction error minimization (den Ouden, Friston et al. 2009; den Ouden, Daunizeau et al.

2010), repetition suppression (Todorovic, van Ede et al. 2011) and mismatch negativity (Garrido, Friston et al. 2008; Garrido, Kilner et al. 2009). Similarly, there are now numerous quantitative computational models, sometimes complemented with psychophysical tests of their predictions, of specific phenomena such as reinforcement learning (Friston, Daunizeau et al. 2009), attention (Feldman and Friston 2010), eye
p. 94 movement in face perception (Friston, Adams et al. 2012) ↳ and occlusion (Adams, Perrinet et al. 2012), and illusion perception (Brown and Friston 2012).

There is a way to turn around the challenge about whether there is evidence for the network. We know that the brain's organization is extremely complex and bewildering, with hundreds of cell types, multitudes of neurotransmitters, and intricate patterns of long and short range neuronal connections. This is evidence against a simple mechanism, re-iterated Lego-block style throughout the brain (Koch 2013). I am hopeful that in time, more and more evidence will come in that will show how this complex organ is in fact engaged in prediction error minimization (for example, in terms of extracting the top-down and bottom-up pattern of activity and their time course in the cortical layers of microcolumns of the visual cortex; perhaps in the style of electrophysiological work using multi-contact transcortical electrodes; for an example of this methodology, see Maier 2008). The challenge can also be met by noting the many aspects of the prediction error scheme, which I have focused upon throughout this first part of the book. Even though the basic idea of prediction error minimization is simple, the implementation of the mechanism is highly complex in a way that calls for many different types of parts and processes: there is a need for first order and second order statistics in the brain's attempt to optimize precisions; there is a need to distinguish perceptual from active inference, and their different directions of fit; there is a need to maintain hypotheses across multiple timescales, calling for different kinds of plasticity; and there is a need to maintain some overall balances, for example between perceptual and active inference. It seems likely that very many different neuronal parts and processes have evolved to maintain these many functions, which enable prediction error minimization and which allow us to fine-tune it in perception, action, and attention.

The other type of challenge flows from the explanatory ambition of the framework. It is meant to apply to all aspects of perception, attention, and action. This means it should be able to provide attractive solutions to recalcitrant problems in cognitive science and philosophy of mind. The challenge in doing so is to avoid just-so stories. That requires avoiding priors and likelihoods that are posited only in order to make them fit an observed phenomenon. To avoid just-so stories any particular ordering of priors and likelihoods should be supported by independent evidence, which would suggest that this ordering holds across domains. Mostly, I shall seek to avoid just-so stories by building specific accounts on just the tools provided by the prediction error minimization mechanism rather than relying on any specific ordering of priors and likelihoods.

Explanations can be attractive for many different reasons. Perhaps they explain a lot of evidence, or key evidence, or they can explain some evidence with great precision, or they can be fecund in terms of
p. 95 generating new research questions, or they are simple, or they are unifying, or integrate well with prior ↳ knowledge, and so on. So often one must weigh up such explanatory virtues when deciding which explanation is best. It is not always obvious which assessment policy is best, and different contexts and interests may call for different approaches and trade-off between virtues. But in general, one should be able to do this kind of weighing of explanatory "bestness", and then engage inference to the best explanation (for an excellent treatment of inference to the best explanation, see Lipton 2004). Inference to the best explanation is capturable in Bayesian terms: namely in terms of which hypothesis best explains away the evidence. So Bayes' rule applies to the Bayesian brain hypothesis. Indeed, like natural selection, the principle of free energy minimization can be applied to itself. In this meta-theoretical approach, models and hypotheses about ourselves should minimize free energy through a judicious balance between maximizing accuracy (explaining empirical observations) and minimizing complexity (thereby providing a parsimonious and unified account of the facts).

The concrete challenge is to pick the right assessment of the explanatory project pursued in this book. Clearly, the project cannot be expected to excel at all explanatory virtues. Personally, I am attracted to the framework because it appears to explain key, philosophical issues mechanistically and as unified under one neuronal mechanism. The explanation is sometimes mechanistic, showing how a phenomenon arises from the properties of the mechanism, and sometimes the explanation is unifying, redescribing a phenomenon to bring it into the prediction error fold, and showing how it connects to other phenomena. Of course, I do not provide much in terms of alternative explanations here. Part of the reason is that I don't know any other theory that can as much as begin to solve the problem of perception—competing theories do not appear to pass the very basic "bestness" hurdle for candidates for inference to the best explanation, namely that of being explanations in the first place. The burden of the book is then to demonstrate that the prediction error minimization scheme is not merely the best out of a poor bunch of explanations, but is good in its own right.

Summary: tooling up for understanding the mind

Now we can begin to form a systematic picture of the nature of the prediction error minimization mechanisms that is iterated throughout the brain. This allows us to "tool up" for the project of Part II of this book, where the framework presented in Part I is put to work.

The mind works by minimizing prediction error. This explains perceptual inference and, in the shape of
p. 96 active inference, action too. The emerging ↳ picture is that the organism needs to alternate between perceptual and active inference; it needs to assess precisions optimally so as to efficiently balance sensory sampling against reliance on prior belief; it needs to balance the complexity of its models against their accuracy and usefulness in active inference. All of these processes involve context-dependence and depend on prior learning. Architecturally, generative models are maintained in a perceptual hierarchy that makes sense of our first-person perspective and which reconceives sensory input as, functionally speaking, a feedback signal on the predictions based on the generative model.

As a whole, this presents the sophisticated implementation of the simple prediction error minimization idea. It delivers a complex, layered functional role for perception and action, for which neuroscience can provide the realization in terms of hierarchical, interconnected patterns of synaptic connectivity, synaptic plasticity, and synaptic gain.

There are weighty reasons to find this framework attractive. It has great unificatory power, it promises to explain in detail, it can be modelled in precise mathematical terms, and evidence is coming in that the brain is in fact working this way, though much empirical work needs to be done. From a philosophical point of view it is attractive because it offers to explain perception, action, and attention on the basis of a naturalistic mechanism that seems up to the task. There are also, as we just saw, challenges to respond to and pitfalls to avoid.

With these few but powerful tools it is possible to re-asses, recalibrate, and reconceive many issues in philosophy of mind and cognitive science. In some cases this allows us to see new kinds of solutions to long-standing problems.

Notes

Page 75. ["At the end of the last . . ."] Here I say that we use perception to guide action. This is a commonplace but is in fact controversial because there is discussion of the extent to which we use *conscious* perception to guide action. Visual illusions for example have surprisingly little effect on movement (Aglioti, DeSouza et al. 1995). I discuss this a little further in Chapter 7.

Page 76. ["Perceptual inference has been presented…"] Formally, mutual information is symmetric so that the mutual information between $P(a)$ and $P(b)$ is the same as the mutual information between $P(b)$ and $P(a)$. Mutual information is the KL divergence between $P(a, b)$ and $P(a)P(b)$, and whereas the KL divergence is not in general symmetric, it is in this instance (see notes to Chapter 2).

Page 77. ["Indeed, it falls natural to look for a role for action in perception…"] Here I focus on the active element in perception. This idea has been taken up and made central to an understanding of perception under the heading of enactive perception, and related to embodied, situated cognition. This is sometimes taken to be a severe ↳ challenge to the idea of internal representations on which prediction error minimization is based (O'Regan and Noë 2001; Noë 2004). It seems clear however that the prediction error approach can encompass the active element of perception without relinquishing internal representations. Conversely, though there is much to admire in the enactive approach, I think it is doubtful it can ultimately do without just the kinds of prediction-generating internal models described here (see, e.g., Block 2005).

p. 97

Page 77. ["Another reason to focus…"] The quote from Helmholtz is my translation. The original in full is "Wir überlassen uns nämlich nicht nur passiv den auf uns eindringenden Eindrücken, sondern wir *beobachten*, das heisst wir bringen unsere Organe in diejenigen Bedingungen, unter denen sie die Eindrücke am genausten unterscheiden können".

Page 82. ["Before moving on…"] Predictions of the flow of sensory states occur as prior beliefs in the prediction error minimization scheme. They play the role of control functions in optimal control theory but when interpreted as priors they can be treated in the same way as perception (Friston, Samothrakis et al. 2012). There are many further aspects to learning and shaping these expectations of flow, including exploratory elements, learning from action behaviour and ideas such as that some goal states are obviated in a way that triggers renewed exploration and action (Friston, Daunizeau et al. 2010).

Page 90. ["Balancing perceptual inference and active…"] The importance of oscillating between perception and action is also discussed by Metzinger (2004: Ch. 6–7), within his framework of theoretical and practical 'phenomenal models of the intentionality relation'.

Page 91. ["We should therefore aim to alternate . . ."] Here I describe the importance of not getting stuck with one active or perceptual inference. This is consistent with recent work on autovitiation of the states visited by self-organising systems, in other words it relates to "a delicate balance of dynamical stability and instability" (Friston, Breakspear et al. 2012: 2).

Page 92. ["There will be an intricate…"] The notion of complexity reduction has given rise to an interesting approach to sleep and dreaming as essentially the brain taking the opportunity to engage in complexity reduction during periods where the light tends to be bad so that the precision of visual input diminishes the reliability of normal prediction error minimization through perception and action (Hobson and Friston 2012).

Page 92. ["Our current state of mind…"] This view of medium level fineness of perceptual grain seems to me compatible with proposals concerning consciousness on which experience is a kind of middle level affair, neither too high nor too low in the perceptual hierarchy; see discussion in (Jackendoff 1987; Koch 2004; Prinz 2012).