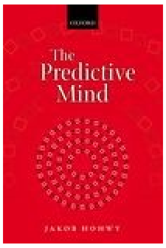


**The Predictive Mind**

Jakob Hohwy

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

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



CHAPTER

# 1 Perception as causal inference

Jakob Hohwy

https://doi.org/10.1093/acprof:oso/9780199682737.003.0002 Pages 12–40

**Published:** November 2013

## Abstract

Our senses are bombarded with input from things in the world. On the basis of that input, we perceive what is out there. The problem that is the focus here is how the brain accomplishes this feat of perception. This chapter pursues the idea that the brain must use inference to perceive — the brain is a Bayesian inference mechanism. The first aim is to show why we should agree with this and what the key ingredients of such perceptual inference are. The second aim is to show how inference could underpin the phenomenology of perception. The chapter describes how perceptual inference is embedded in a perceptual hierarchy of increasing time scales, maintained in the brain. This gives reason to believe that perceptual inference can accommodate the richness of perceptual phenomenology, as perception encompasses both variant and invariant representation; much of this is illustrated with examples from perceptual science, in particular binocular rivalry. The chapter ends with a brief primer on Bayes' rule.

**Keywords:** problem of perception, Bayesian perceptual inference, hierarchical perceptual inference, binocular rivalry, Bayes' rule

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

**Collection:** Oxford Scholarship Online

Our senses are bombarded with input from things in the world. On the basis of that input, we perceive what is out there. The problem that will concern us is how the brain accomplishes this feat of perception.

This chapter pursues the idea that the brain must use inference to perceive—the brain is an inference mechanism. The first aim is to show why we should agree with this and what the key ingredients of such perceptual inference are. The second aim is to show how inference could underpin the phenomenology of perception.

A very basic and useful formulation of the problem of perception is in terms of cause and effect. States of affairs in the world have effects on the brain—objects and processes in the world are the causes of the sensory input. The problem of perception is the problem of using the effects—that is, the sensory data that is all the brain has access to—to figure out the causes. It is then a problem of causal inference for the brain, analogous in many respects to our everyday reasoning about cause and effect, and to scientific methods of causal inference.

The problem of perception is a *problem* because it is not easy to reason from only the known effects back to their hidden causes. This is because the same cause can give rise to very different effects on our sensory organs. Consider the very different inputs we get from seeing rather than merely touching a bicycle, or seeing it from different perspectives, or seeing it in full view as opposed to being partly obscured behind a bush. Likewise, different causes can give rise to very similar effects on our sense organs. Consider the potentially identical sensory input from different objects such as a bicycle or a mere picture of a bicycle, or a whole bicycle occluded by a bush as opposed to detached bicycle parts strewn around a bush, or more outré possibilities such as it being an unusually well-coordinated swarm of bees causing the sensory impression as of a bicycle.

In our complex world, there is not a one-one relation between causes and effects, different causes can cause the same kind of effect, and the same cause can cause different kinds of effect. This makes it difficult for the
p. 14 brain to pick the one effect (sensory input) that goes with the one cause (object in the ↳ world). If the only constraint on the brain's causal inference is the immediate sensory input, then, from the point of view of the brain, any causal inference is as good as any other. When the input is different, as in the seen and felt bicycle case, the brain would not know whether to infer that the cause of the inputs is the same, or if there are distinct causes, and whether one type of cause is more likely than another.

## Constraints on perceptual inference

The key issue is then that without any *additional constraints* the brain will not be able to perform reliable causal inference about its sensory input. We can in fact engage in such inference, since we can perceive. So there must be such additional constraints, but what could they be?

One possibility is that the additional constraints are *mere biases*. Even though the brain cannot reliably infer that it is one rather than another cause, it simply happens to be biased in favour of one. It just so happens that it decides in favour of, say, the bicycle being the cause when it gets a certain kind of input. No doubt there is a describable, law-like regularity in nature such that, in certain to-be-specified conditions, if a system like the brain were to have a certain kind of sensory input caused by a bicycle, then it would be biased towards perceiving it as a bicycle. In principle, various branches of science would be able to discover these biases by systematically exposing systems like the brain to bicycle inputs and tracking the causal chain of events throughout the brain. The brain would seem to cut through the problem of perception by just opportunistically favouring one among the intractably many possible relations between cause and effect.

But even if at some level of description there are these regularities it would not solve the problem of perception as we have conceived it. Such regularities do not afford an understanding of perception as causal *inference*. Inference is a normative notion and brute biases cannot lead us to understand how there could be a difference in quality between an inference back to bicycles rather than, say, swarming bees being the cause of sensory input. What brute regularities in nature give us is a story about what the system *would* do, not what it *should* do in order to get the world right. What is needed, then, is a normative understanding of the role of such regularities. We need to see the additional constraints on causal inference in normative terms.

There is a clear first candidate for an additional constraint with normative impact. It seems obvious that causal inference about things like bicycles draws on a vast repertoire of *prior belief*. This could be what allows us to rank lowly some candidate causes such as it being a swarm of bees that is causing the current

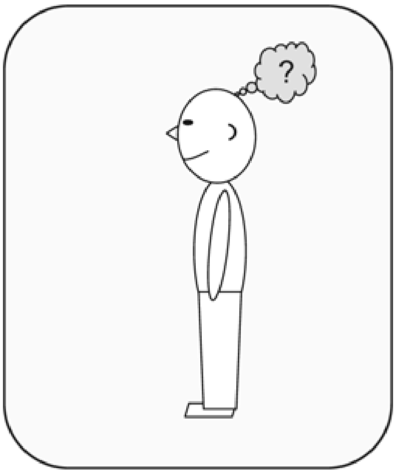
p. 15 sensory impression. Our prior experience tells us that bees are actually ↳ extremely unlikely to form such patterns of sensory input in us. There is in fact little doubt that perceptual causal inference needs to be buttressed with prior knowledge, but doing so is no trivial matter. On the one hand, if the story we tell is that we just find ourselves with a stock of prior beliefs, then we have not after all moved beyond the mere biases type of story. On the other hand, if prior knowledge is itself a product of prior perceptual, causal inference, then we are presupposing what we set out to explain, namely perceptual causal inference—the bump in the carpet has merely shifted.

We can now see what a solution to the problem of perception must do. It must have a bootstrapping effect such that perceptual inference and prior belief is explained, and explained as being normative, in one fell swoop, without helping ourselves to the answer by going beyond the perspective of the skull-bound brain (Eliasmith 2000; Eliasmith 2005). The contours of just such a solution are now beginning to emerge. It is based in probability theory—Bayesian epistemology—which is normative because it tells us something about what we should infer, given our evidence.

## Perception and Bayes' rule

Consider this very simple scenario. You are in a house with no windows and no books or internet. You hear a tapping sound and need to figure out what is causing it (Figure 1).

**Figure 1.**




The basic perceptual inference problem: figuring out what caused a sound. This is analogous to the situation for the brain.

This illustrates the basic perceptual task. You are like the brain, the house is the skull, and the sound is
p. 16 auditory sensory input. As you are wondering about ↳ the cause of the input, you begin to list the possible causes of the input. It could be a woodpecker pecking at the wall, a branch tapping at the wall in the wind, a burglar tampering with a lock, heavy roadworks further down the street, a neighbour's loud music, or those kids throwing stones; or it could be something internal such as loose water pipes banging against each other. Let your imagination rip: it could be that your house has been launched into space over night and the sound is produced by a shower of meteorites. There is no end to the possible causes. Call each of these possibilities a *hypothesis*. The problem of perception is how the right hypothesis about the world is shaped and selected.

Set aside the problem that once we begin generating hypotheses, there is no clear principle for when we should stop. Consider instead the fact that we *can* generate hypotheses, and that not just any hypothesis will seem relevant. For example, we would not accept that the tapping noise on your house could be produced by a distant mathematician's musings on Goldbach's conjecture, or by yesterday's weather. This means we are able to appreciate the link between a hypothesis and the effects in question. We can say "if it is really a woodpecker, then it would indeed cause this kind of sound". We can say something about how likely it is that the hypothesis fits the effects. This is *likelihood*: the probability that the causes described in the hypothesis would cause those effects. It is clear that assessing such likelihoods is based on assumptions of causal regularities in the world (for example, the typical effects of woodpeckers). Based on our knowledge of causal regularities in the world we can often rank hypotheses according to their likelihood, according to how close their tie is to the effects we are seeking to explain. Such a ranking can be said to capture how good the hypothesis is at accounting for, or *predicting*, the effects. For example, the woodpecker hypothesis may have roughly the same likelihood as the banging pipes hypothesis, and both have higher likelihood than the hypothesis concerning those stone-throwing kids.

We could simplify the problem of perception by constraining ourselves to just considering hypotheses with a high likelihood. There will still be a very large number of hypotheses with a high likelihood simply because, as we discussed before, very many things could in principle cause the effects in question. Just going by the hypothesis with the very highest likelihood does not ensure good causal inference. Here is a hypothesis with very high likelihood: the sound is caused by a tapping machine especially designed by cunning neuroscientists to use you to illustrate perceptual causal inference. This hypothesis fits the auditory evidence extremely well, but it does not seem like a good explanation in very many actual situations. The problem is that the cunning neuroscientist hypothesis seems very improbable when considered in its own right and before you heard the banging sound.

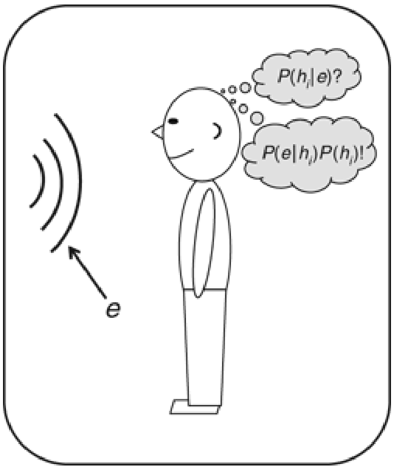
Therefore, we need to take the independent, prior plausibility of hypotheses into consideration, in addition to their likelihood. We need to consider the probability of the hypothesis prior to any consideration of its fit
p. 17 with the ↳ evidence. This is then the *prior probability* of the hypothesis. Perhaps there is some objective truth about how probable each hypothesis is, based on the frequency of the events it describes. This kind of knowledge would be useful but mostly it is not something we have. Instead we will assume you assign probabilities to hypotheses based on your own background beliefs and subjective estimates (making sure the probabilities sum to 1, to make the ranking meaningful).

By appealing to your prior beliefs we have given you two tools for figuring out the cause of the sound: likelihood, which is the probability of the effect you observe in the house *given* the particular hypothesis you are considering right now; and the prior probability of the hypothesis (or just the "prior"), which is your subjective estimate of how probable that hypothesis is independently of the effects you are currently observing.

It seems rational to pick the hypothesis which best fits the observed effects but weighted by the independent probability of that hypothesis. Likelihood and prior are the main ingredients in Bayes' rule, which is a theorem of probability theory and thought by many to be a paradigm of rationality. This rule tells us to update the probability of a given hypothesis (such as the woodpecker hypothesis), given some evidence (such as hearing some tapping sound) by considering the product of the likelihood (which was the probability of the evidence given the hypothesis) and the prior probability of the hypothesis (normalized so probabilities sum to 1). The resulting assignment of probability to the hypothesis is known as the *posterior probability*. The best inference is then to the hypothesis with the highest posterior probability. (A brief primer on Bayes' rule is included at the end of this chapter).

Return now to the sound you hear in the house. With likelihoods and priors you can arrive at a good hypothesis: the one that achieves the highest posterior. If you have experienced many woodpeckers in your area and only a few burglars, and if you don't really think your house is likely to have been launched into space over night, and so on and so forth, then you should end up inferring to the woodpecker hypothesis (Figure 2).

**Figure 2.**




Prior probability of hypothesis $h_i$: $P(h_i)$. Likelihood that the evidence e would occur, given $h_i$ is the true hypothesis: $P(e|h_i)$. Posterior probability of the hypothesis $h_i$, given the evidence $e$: $P(h_i|e)$. Simplified version of Bayes' rule that puts it together: $P(h_i|e) = P(e|h_i)P(h_i)$.

Even on this very simplified presentation, Bayesian inference provides a very natural way to think about perception. Of course, the drawback with illustrating the problem as I have done here is that there is no intelligent little person inside the skull consciously performing causal inference. On the story we shall develop, which goes back to Helmholtz, what is really going on is that the neural machinery performs perceptual inference unconsciously. As Helmholtz says about the "psychical activities" leading to perception,

> [they] are in general not conscious, but rather unconscious. In their outcomes they are like inferences insofar as we from the observed effect on our senses arrive at an idea of the cause of this effect. This is so even though we always in fact only have direct access to the events at the nerves, that is, we sense the effects, never the external objects (Helmholtz 1867: 430).

p. 18 So what we will be talking about is *unconscious perceptual inference*. The job before us is to see how what the system does can be usefully conceived as a form of inference. We just need to accept the Helmholtzian idea that the brain is capable of unconsciously going through the same kind of reasoning that we described for figuring out the cause of the sound heard inside the locked house. The brain infers the causes of its sensory input using Bayes' rule—that is the way it perceives. The core idea is fairly clear and has a pleasing air of generality to it: the problem of perception is not inherently special, something for which an entirely new branch of science is needed. It is, instead, nothing more than a version of the kind of causal inference problem that we are often confronted with in both science and everyday life.

While the Bayesian, inferential approach to perception is attractive many questions quickly arise. Straight off, aligning perception with ideally rational, probabilistic, scientific-style reasoning seems rather intellectualist. It is difficult to learn probability theory and to implement Bayesian inference but perception is unconscious and effortless—it is something adults, children, and animals can do without knowing anything about Bayes. Moreover, there is evidence that we are not very good at explicit Bayesian reasoning —Bayes' rule takes some explaining and exercise so does not seem to come naturally to us (Kahneman, Slovic et al. 1982). There is also something slightly odd about saying that the brain "infers", or "believes" things. In what sense does the brain know Bayes, if we don't?

For that matter, a Bayesian approach to perception does not seem to directly concern the full richness of perceptual phenomenology as much as mere conceptual labelling or categorization of causes of input (it p. 19 could seem to be not so ↳ much about visually experiencing a bicycle as merely labelling some sensory input "bicycle"). Nor does this approach, with its focus on assigning subjective probabilities, immediately begin to provide a satisfactory explanation of where prior beliefs come from. As we will see in this and the following chapters, the theoretical framework can be developed to deal with all of these issues.

The contrast to the inferential picture of perception is a picture on which perception, rather than being the upshot of inferential processes in a hypothesis-testing brain, is the result of an analytic, bottom-up driven process where signals are recovered from low-level sensory stimulation and gradually put together in coherent percepts. On this alternative, non-inferential approach, perception is driven bottom-up by the features the brain detects in the input it gets from the world. Crudely, changes in input drive changes in perception, and so top-down inference in any substantial, normative sense is not needed.

There is much discussion about the relative virtues of the feature detection approach vs. the more inferentialist, Bayesian approach (for a review and discussion, see Rescorla (in press). One reason for not adopting the feature-detection approach is that it is not very clear how it can help with the problem of perception as we have set it out above. This theoretical debate cannot be resolved conclusively here but in the next section I will give what I think is a very good example of a perceptual effect demonstrating the need for inference.

## Perceptual inference and binocular rivalry

In 1593 the Italian polymath Giambattista della Porta reported an intriguing visual phenomenon:

> Place a partition between the eyes, to divide one from the other, and place a book before the right eye, and read; if another book is placed before the left eye, not only can it not be read, but the pages cannot even be seen, unless the visual virtue is withdrawn from the right eye and changed to the left (Porta 1593; quoted in Wade 1998: 281).

Some centuries later Charles Wheatstone invented the stereoscope, which uses mirrors to help split the images presented to the eyes, and in 1838 also described this kind of perceptual alternation between different letters shown to each eye (Wade 1998; Wade 2005). This fascinating effect is known as binocular rivalry and remains, 400 years after Porta, a vibrant focus of much research in vision science. The neural mechanism behind it is still unknown and it keeps throwing up new and intriguing findings. As Porta delightfully puts it, what makes "visual virtue" alternate between the eyes?

It is a surprising effect because one would think that if two different images are shown to the eyes they p. 20 should just somehow blend in with each ↳ other. If a picture of a house is shown to one eye and a picture of a face is shown to the other, then one should surely just see a face-house. But this is not what happens, as Porta and Wheatstone and many others have described. The brain somehow seems to *decide* that there are two distinct things out there, a face and a house—and perception duly alternates between seeing one or the other every few seconds, sometimes with periods of patchy rivalry in between.

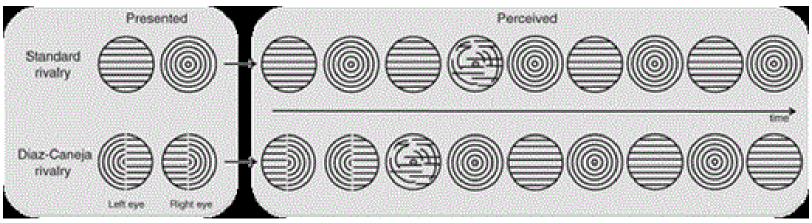
We shall return to binocular rivalry on a number of occasions later in this book but for now notice that it puts pressure on the idea that perception is purely stimulus driven, bottom-up feature detection. During rivalry, the physical stimulus in the world stays the same and yet perception alternates, so the stimulus itself cannot be what drives perception. It is very difficult not to impute to the perceptual system some manner of inferential power here. It is as if the perceptual system refuses to accept that a reasonable solution to a confusing input could be a face-house mishmash out there in the world.

To put it in the Bayesian vernacular, the prior probability of such a mishmash cause of my perceptual input is exceedingly low. Instead, very “revisionary” hypotheses are selected, each of which effectively suppresses a large part of the incoming sensory signal. It is as if when a face is seen the visual system says “it is most probably a face, never mind all the parts of the total input that the face hypothesis cannot explain”; and *vice versa* when perception then alternates and the house is seen. How exactly this inferential process proceeds is a further matter but it is difficult to see how we could even begin to explain this effect without appealing to some kind of inference.

Recall the worry that the Bayesian, inferential approach to perception seems rather intellectualist. The initial response to this is then that some degree of inference seems to be necessary at least in some circumstances. It is of course possible that the brain only has to resort to this kind of inference in special cases like rival input to the eyes. However, it would be odd if the brain had evolved a highly sophisticated inferential process to deal with a perceptual situation it encounters mainly in highly artificial laboratory settings (though there is debate about how uncommon it is, see Arnold 2011; O’Shea 2011). It seems reasonable to work on the assumption that the brain always uses some sort of inferential process to perceive the world, and that rivalry is an effect that simply makes the brain’s everyday inferential processes easier to spot.

Some of the wonderful special aspects of binocular rivalry strengthen the supposition that the brain is engaged in quite sophisticated inferential work. In 1928 Emilio Diaz-Caneja (Diaz-Caneja 1928) discovered that if the two images are cut in half and combined such that one eye sees, for example, half a house and half a face, and the other eye sees the other halves of the house and the face, then there is not rivalry between what is presented to each eye, p. 21 ↳ there is instead rivalry between the full, uncut images of the face and the house (Figure 3 illustrates this for the type of stimuli Diaz-Caneja used).

**Figure 3.**




Top panel: standard rivalry. Bottom panel: rivalry with Diaz-Caneja stimuli; the different halves of the stimuli are grouped together, so that perception often resembles standard rivalry.

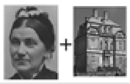
This is a remarkable feat by the brain. It is also a stunning thing to experience for oneself. It shows that even if rivalry is to some extent is the result of very low-level brute competition between processing from each eye this cannot be the whole story since half an image is taken from each eye and grouped in coherent, rivalling percepts.

In the same vein, work in the lab of leading neuroscientist Nikos Logothetis has demonstrated that if the images to the eyes are swapped from eye to eye a couple of times every second, rivalry continues in a relatively normal fashion. So, if you’re currently seeing the face, as shown to the right eye, then you will continue to see the face, even if the image of the face in the right eye is swapped to an image of the house (Logothetis, Leopold et al. 1996). The brain very dramatically overrules the actual input in order to make sense of the world.

With Andreas Roepstorff and Karl Friston I have suggested a simple Bayesian story about why only one image is seen at a time in binocular rivalry. The visual system receives an unusual total input in the shape of, for example, a house image to one eye and a face image to the other. There are three relevant, candidate

hypotheses to explain what may have caused this sensory input: it is a house only, it is a face only, or it is a face-house blend. The system will pick one of these hypotheses based on (i) their likelihood, that is, how likely it is that a house, face, or face-house blend would have caused this input, and (ii) their prior probability, that is, how likely it is that you should be seeing a house, a face, or a face-house now, irrespective of the actual sensory input. The Bayesian story then goes like this. The combined face-house blend hypothesis has the highest likelihood, because it accounts for more of the sensory input than the face or the house hypotheses on their own. But this high likelihood cannot overcome the exceedingly low probability that a face and a house could co-exist in the same spatiotemporal location (you might on
p. 22 occasion come across a transparent image of a face positioned in front of a house but it is ↳ very difficult to conceive of fully opaque faces and houses in the very same location in space). So the hypothesis that is selected, and which determines perception, is either the face or the house hypothesis; Figure 4 (Hohwy, Roepstorff et al. 2008).

**Figure 4.**

| | |
|---|---|
| Input: I |  |
| Hypotheses | F+H: "It's a face-house"<br>H: "It's a house"<br>F: "It's a face" |
| Likelihoods | P(I/F) = P(I/H) < P(I/F+H) |
| Priors | P(F) > P(H) >> P(F+H) |
| Perceptual inference | P(F/I) > P(H/I) > P(F+H/I)<br> |

The selection of one of two presented stimuli for perception in binocular rivalry, interpreted in Bayesian terms (adapted from Hohwy, Roepstorff et al.2008).

Some empirical support for this hypothesis is emerging. If the Bayesian story is correct it follows that, if the prior probability for one of the hypotheses goes up, then that enhanced hypothesis should dominate in rivalry. Rachel Denison and colleagues (Denison, Piazza et al. 2011) used lines at different orientations to each eye to induce rivalry and successfully biased prior probabilities in favour of one or the other eye's stimulus by briefly showing rotating lines that would stop at either the horizontal or vertical position just before rivalry would begin. As predicted, participants are more likely to select for their first percept the stimulus with the highest prior probability. Zhou Wen and colleagues (Zhou, Jiang et al. 2010) induced binocular rivalry by presenting participants with images of text markers and roses. They increased the probability of it being the roses by adding olfactory evidence and letting participants smell roses too. As predicted by the Bayesian story, the participants consequently spent more time perceiving the rose image.

As we will see later, more needs to be done on this simple Bayesian account. For one thing, it does not explain why there is continued *alternation* between images in binocular rivalry (Chapter 10 returns to this issue, and somewhat revises the proposal given in Hohwy et al. 2008). As presented so far, the account only
p. 23 explains how it can be that only one image is selected for ↳ perception. Nonetheless the basic Bayesian idea at least begins to make sense of some essential features of rivalry.

The jury is still out about how much the Bayesian approach can contribute to the understanding of binocular rivalry and sensory processing in general (for a review, see Blake and Wilson 2011). While I am not suggesting the appeal to rivalry will end this debate once and for all, it strongly suggests that even if none of us consciously know and apply Bayes' rule in perception, the perceptual system in our brains somehow unconsciously follows Bayes' rule. What the brain does for us is indeed inferential. I think binocular rivalry provides a particularly good case in defence of perceptual inference, though Hermann von Helmholtz and many others since have appealed to a number of additional phenomena in their defence of the notion of unconscious perceptual inference.

## How do neurons know Bayes?

The proposal is that the brain unbeknownst to consciousness is engaged in sophisticated probabilistic reasoning. This may sound as if neuronal populations in the visual cortex and throughout the brain know and apply Bayes' rule. Putting things like this carries a risk of what we might call neuroanthropomorphism —inappropriately imputing human-like properties to the brain and thereby confusing personal level explanations with subpersonal level explanations. As we saw, there are strong reasons to think perception is unconscious inference, carried out by the brain, so the question is how we should understand this idea without succumbing to crass neuroanthropomorphism.

A huge theoretical issue lies buried here, which we will not resolve fully. But since I will continue to talk in terms of the brain "inferring", "believing", and "deciding" things it may be useful to briefly explain why I do not think this usage is particularly problematic. Hopefully, in the balance of the book, it will then come to seem natural.

An analogy to the problem comes from computer science and research in Artificial Intelligence: the components of a computer chip do not in any ordinary sense "know" the concepts of the program they are executing. Accordingly, there is a theoretical debate about how software really stands to hardware. This kind of debate plays out in philosophy of mind too, where *functionalists* about mental states hold that mental states are defined by a functional role that specifies a certain kind of input-output profile, given a certain internal state. A toy example: you are in the mental state of pain if you have an input of bodily damage and an output of screaming and pulling your hand from the fire, given you're in the internal state of desiring to
p. 24 avoid bodily ↳ damage and believing that moving your hand will help. Functionalists then discuss the relation between this functional role and the physical stuff that plays the role, which in us is an extended network in the brain often labelled the pain matrix.

Various terms such as "implementation" or "realization" can be used for the relation between the role and what plays the role, as well as for the relation between computer program and computer chip (I will tend to use "realization" in this book). There is debate about what these notions mean exactly but for our purposes this laxness is not critical. There is also debate about the extent to which functionalism is independent of neurobiological detail. Some functionalists argue that it is the functional role and not its realization that is crucial to understand a phenomenon, others insist the realization is crucial. There is further debate about how computational approaches relate to functional roles, and to neurobiological mechanisms (for a nice discussion that resolves this debate, see Kaplan 2011). Even though there are heavy functional considerations behind the framework that is discussed in this book, there is also a direct motivation from the neurobiological detail of the brain's structure, namely in terms of its hierarchical structure and the nature of message passing within the brain; this strongly suggests a Bayesian brain, implemented with prediction error minimization, as we will see in the next chapter.

Very few people would claim that computers do not engage in computation because the hardware inside them does not *know* the concepts and rules employed in the program. Similarly, we should not claim that brains do not engage in probabilistic inference because the neurons making them up do not know Bayes' rule. What we should claim, rather, is that we can only understand how computers engage in computation if we understand how the hardware is able to realize the functional roles set out in computer programs. Similarly, we can only understand how brains engage in probabilistic inference if we understand how neurons can realize the functional roles set out by forms of Bayes' rule. Getting to fully understand this is not going to be a trivial task but putting it like this somewhat defuses the worry that the Bayesian approach to perception is crudely neuroanthropomorphic: if it was, then so too would be the claim that computers compute.

Of course, this is a fairly brief defence of the application of Bayesian vernacular to the brain. Underlying it is a more substantial view based on the rather uncontroversial idea that the brain is involved in information processing, and that information theory is cast in terms of the probability theory from which Bayes' rule is derived (see the last section of this chapter for a primer on this derivation). It would be odd, therefore, if the brain's processing could not be understood in ways that could at some level of description smoothly involve Bayes' rule. This kind of sentiment is captured well by Chris Eliasmith in an argument that our conception p. 25 of the mind is ready to move ↳ beyond metaphors of symbol manipulation, connectionism, and dynamics: "We are in the position, I think, to understand the mind for what it is: the result of the dynamics of a complex, physical, information processing system, namely the brain" (Eliasmith 2003: 494).

In many ways, this broad line of reasoning is the impetus for this book: there is converging evidence that the brain is a Bayesian mechanism. This evidence comes from our conception of perception, from empirical studies of perception and cognition, from computational theory, from epistemology, and increasingly from neuroanatomy and neuroimaging. The best explanation of the occurrence of this evidence is that the brain *is* a Bayesian mechanism. So, by inference to the best explanation, it is. I find the appeal to inference to the best explanation attractive as a way through this debate—not least because this inference type is itself essentially Bayesian.

## From inference to phenomenology

So far, this chapter has built up the case in favour of using unconscious probabilistic inference as an approach to perception. Now it is time to consider how such rather austere looking inference can construct the richness of perceptual experience.

I used binocular rivalry to illustrate the need for a notion of inference in perception. Rivalry can also be used to start the discussion about perceptual phenomenology. What happens in rivalry is not that what one sees is an unchanging, confusing mishmash of the two pictures while one's conceptual judgement alternates—it is not that you will *see* a face-house blend and *think* "it's a house…no, it is a face…wait, no, it is a house…". What makes rivalry so intriguing is that what changes is what you actually see, that is, the inferential process drives perceptual content itself.

My own first experience of rivalry was when doing a kitchen table experiment. I placed a blue and a red toy car (a Porsche and a van) on the table, viewing each through a cardboard toilet paper roll and trying to free-fuse my eyes to make the cars appear in the same location in the visual field. After some trial and error it worked and took my breath away. There are periods of patchy rivalry where I see bits of blue car and bits of red car, as if the three hypotheses of what is out there in the world (mentioned in the previous chapter, see Figure 4) are fighting each other. Then one of the blue patches begins to spread and suddenly I see only the blue Porsche, nothing is left of the red van even though I know full well it is presented to one of my eyes. A few seconds later, a red corner of the van pops up and spreads to suppress any hint of the blue car.

p. 26 Rivalry is characterized by this very dramatic change in actual visual consciousness. It is as if the brain uses vibrant mental paint to paint over the image from one eye. What you see may cause you to issue a conceptual judgement such as thinking "this is a blue Porsche" but it is the visual perception itself that is changing so dramatically in rivalry. In the words of Helmholtz, rivalry is a "wonderful theatre" ("ein wunderliches Schauspiel") (Helmholtz 1867: 776). It is that perceptual phenomenology that we are trying to explain in terms of probabilistic inference.

Straight off, it can seem that the notion of perceptual inference, in its simple Bayesian clothing, is just a labelling exercise that may allow us to recognize or categorize objects. It is more difficult to cast it as a process that can yield the rich perceptual content of actually seeing a blue Porsche, or all the features of a face. There only seems to be competition between labels for what is seen ("should I categorize this input as 'blue Porsche', 'red van', or 'blue-red Porsche-van'?"). So the task now is this: show that the inferential approach to perception can accommodate differences in perception as such, rather than only differences in conceptual categorization.

This is an important task, and central to the message of this book. It seems the Bayesian moves we just made would do fine if we were only interested in explaining conceptual thought about perception, rather than explaining perception itself. So what makes this an account about perception specifically?

There is an answer to this question. To see this it is necessary to appreciate a *hierarchical* notion of perception. Bayesian perceptual inference applies to all levels of sensory attributes, and perception normally simultaneously takes in a wide range of these levels. These levels of sensory processing are ordered hierarchically and this is a crucial aspect of the account of the hypothesis-testing brain.

Specifically, this hierarchical notion of perceptual inference seems able to capture something central about perceptual experience, which sets it apart from mere categorization or labelling, namely that perception is always from a *first-person perspective.* It is not just that we see a car but that we see it, as a car, from our own perspective. Different levels of our perspectival experience change in concert as the movement of eyes, head, or body changes our perspective on the world. Perceptual content is embedded in the cortical, perceptual hierarchy and there can be dramatic changes in this content as our first-person perspective changes. This teaches us that what we say about how things are, how we end up categorizing them, depends on how things more transiently seem to us in during perspectival changes. I now set out the notion of a perceptual hierarchy and attempt to capture these aspects of our perceptual phenomenology.

p. 27

## A hierarchy of causal regularities

The world is rife with regularities. Day follows night, seasons follow one another, most power corrupts, milk goes sour, faulty brakes are often followed by accidents, many marriages are followed by divorce, and so on. These regularities are of a causal nature: faulty brakes cause accidents, the planet's spinning in the solar system cause the succession of day and night, and lots of hidden causes can contribute to divorce. There is also irregularity, or noise. The milk goes sour but there is some variability in exactly when it happens, power corrupts but how much can be difficult to say. Even in the best of circumstances we have to accept a level of irreducible noise. Perception requires us to extract the regularities from the irregularities, the signal from the noise. In science this normally happens by controlling for interfering factors in the lab and intervening judiciously in the causal chain. In normal perception it mostly happens by keeping track of and modelling relevant interfering factors (and with the help of action, attention, and other tricks, as we will discuss in Chapter 4 and 7).

Regularities come at different time scales, ranging from tens of milliseconds to hundreds, to seconds, minutes, and upwards towards regularities or rules that are stable over weeks, months, and years. Fast time-scale regularities include things like how shadows change as you move an object in your hands, slower ones concern the trajectory of a balloon you're trying to catch, slower still concern the way people tend to respond to your requests, and still slower how people tend to vote in years of financial unrest.

Mostly, there is a trade-off between time scale and level of detail. Fast changing regularities are good for detail; slower regularities are more general and abstract. This makes sense when we consider what regularities allow us to predict. If I want to predict something with great perceptual precision then I cannot do it very far into the future, so I need to rely on a fast changing regularity (exit polls are better estimates of voter behaviour than polls a week before the election but allows less time for action). On the other hand, predictions further into the future come at a loss of precision and often detail (on our current form curve I may predict that we will lose next Sunday's game against the Vikings but I will only be able to predict exactly by how much we will lose a few seconds before the final whistle). The relation is intricate because there may well be longer-term regularities about patterns in detailed behaviour. For example, I can predict that every year in September newspapers in Melbourne will be full of words about Australian Rules Football, even if I don't know exactly what those words will be.

Regularities can be ordered hierarchically, from faster to slower. Levels in the hierarchy can be connected such that certain slow regularities, at higher levels, pertain to relevant lower level, faster regularities (for
p. 28 example, slow ↳ regularities about aussie rules footy word frequency during the yearly news cycle pertain to faster regularities about the words I end up reading; if I know the slower regularity then I am less surprised by the occurrence of those words). A complete such hierarchy would reveal the causal structure and depth of the world—the way causes interact and nest with each other across spatiotemporal scales.

Causal structure and depth is important to perception in at least three ways. Causal interactions are what make perceptual inference difficult by preventing simple one-one relations between the causes of my sensory input and the sensory input itself, discussed earlier in this chapter. Causal interactions between objects, and between the perceiver and objects, shape our first person perspectival experience (for example, the way a shadow may disappear and reveal the true shape of an object when we hold it out in the sunlight). Finally, causal structure allows us to plan our own causal interactions with the world on the basis of what we perceive.

The brain responds to this importance of causal, hierarchical structure in a very comprehensive manner: it recapitulates the interconnected hierarchy in a model maintained in the cortical hierarchy of the brain. Fast regularities are processed early in the sensory processing stream (for visual perception, this happens in area V1 at the back of the brain) and then increasing time scales are processed as the sensory signal works its way up through the primary sensory areas and into higher areas.

The hierarchy also has a spatial dimension, which sits naturally with the temporal focus we have had so far. The fast time scale regularities represented in low levels of the hierarchy (such as in V1) have small, detail-focused receptive fields of only a couple degrees whereas later areas of processing have wider receptive fields (e.g., 20–50 degrees in the temporal cortex). Receptive fields are also characterized by interconnections, such that wide receptive fields take in sets of smaller receptive fields processed lower down in the hierarchy.

Perceptual inference happens in this highly interconnected, cortical hierarchy and can as such avail itself directly of its representation of myriad causal relations in its attempt to get the world right, in its construction of a first-person perspective, and in its ability to orient itself for action in the world (Friston 2008; Kiebel, Daunizeau et al. 2008). I will first explore some of the properties of this *perceptual hierarchy* and then, in the next chapter, explain how it is thought to arise in the brain, and how it is shaped.

## Perceptual variance and invariance

Fast regularities occur in perceptual inference in the shape of the *variant* aspect of experience: perception
p. 29 captures our immediate and constantly ↳ varying first-person perspective. Every time there is a difference in first-person perspective, for example as your eyes or head move or objects of perception shift around, the brain needs to process fast causal regularities for very basic sensory attributes such as contour, shading, and orientation. Some of these changes are suppressed, such as those arising from quick saccadic movement of the eyes. But many changes are consciously experienced, such as those caused by moving your head to scan a scene in front of you.

At the same time, slow regularities occur in perception in the shape of the *invariant* aspect of experience: perception depends on our ability to abstract from our immediate fluctuating first-person perspective and focus on states of the world that are less sensitive to the concrete way the world is being sampled by the senses right now. For example, even though there are dramatic differences in fast regularities as you perceive a child playing a basketball match, you perceive an enduring object throughout the game rather than just a rapidly changing series of jumbled, perspectival scenes. As Edmond Rolls, who is the architect of an impressive computational model of invariant object recognition, puts it concerning visual perception:

> One of the major problems that is solved by the visual system in the cerebral cortex is the building of a representation of visual information which allows object and face recognition to occur relatively independently of size, contrast, spatial-frequency, position on the retina, angle of view, lighting, etc. These invariant representations of objects, provided by the inferior temporal visual cortex are extremely important for the operation of many other systems in the brain, for if there is an invariant representation, it is possible to learn on a single trial about reward/punishment associations of the object, the place where that object is located, and whether the object has been seen recently, and then to correctly generalize to other views, *etc.* of the same object. (Rolls 2012: 1)

The difference between variant and invariant perception, as defined here, is best conceived as a matter of degree, with somewhat vague endpoints. As perception becomes more and more dependent on slower regularities it becomes more and more invariant. For example, our perception of people as enduring objects is more invariant than our perception of the change of facial features as a person smiles at us.

It is not entirely clear how fast the time scales are at the variant end of the scale, or exactly how fast the regularities are that we can become conscious of even if we do process them (there is some computational evidence that very basic and fast-changing sensory attributes are dealt with in a Bayesian manner, such as features of line orientations and lengths, see Rao and Ballard 1999). Neither is it clear how slow the regularities are at the invariant end of the scale. For the sake of illustration, a maximally slow regularity could be a Big Bang–Big Crunch cycle of the entire universe but it is very unlikely that this regularity,
p. 30 though we can represent it in various ways, plays any role in ↳ modulating ongoing perceptual inference. On the other hand, the rather constant regularity that light normally comes from above does influence perceptual inference of convexity and concavity (for experiments on this, see Adams, Graf et al. 2004; Morgenstern, Murray et al. 2011). Similarly, the slow regularity that captures how our bodies grow and change as we age is factored into our perception of other people over time, and may modulate our levels of surprise when we see them after a long absence. For example, I am surprised at people who don't seem to age.

It is thus possible to think of degrees of invariance in perception in terms of the spatiotemporal hierarchy of represented causal regularities. This yields a nice conception of your first-person perspective, namely as your actual perceptual inferences ordered and connected according to invariance. Your first-person perspective and mine will differ in as much as we engage in different short time-scale, variant inferences, and they will overlap in as much as we engage in similar inferences over more extended time scales.

This helps explain a key feature of perceptual experience—that it always has a first-person perspective—and thus helps us see how the notion of Bayesian perceptual inference pertains to perception rather than mere object categorization. If Bayesian perceptual inference happens in a recapitulation of the causal hierarchy, spanning a wide spatiotemporal range, then it can encompass both the invariant perception that matters for recognition and planning, and the variance characteristic of the more transient first-person perspective.

The causal hierarchy is thus crucial for a plausible account of perceptual inference. It provides the first step in combining variant (first-person perspectival) and invariant perception within one type of process. This aspect in turn relates to more epistemic issues concerning our ability to know the states of affairs of the world and how we ourselves are placed in it. Sometimes we get to doubt that our perceptual inference is correct because we learn that it depends too much on our variant perspective. That may lead us to reality test, and explore some state of affairs better, more deeply, from different perspectives. The aim of reality testing in such cases is to arrive at more confident perceptual inferences, which are anchored more solidly in invariant perception. Similarly, our perceptual knowledge of states of affairs in the world depends on our personal trajectory through the world and this is something variant perception gives us information about. The flow of variant information allows us to track how we are positioned over against objects in the world. The perceptual hierarchy thus plays a role for how we conceive of our own epistemic role. These epistemic matters belong with the deeper facets of perceptual phenomenology and they seem also to relate to the perceptual hierarchy. I will pursue some of these issues in more detail in Chapter 7.



## Message passing between hierarchical levels

The basic idea for the hierarchy of perceptual inference is that for every level of the hierarchy we probe deeper into the causal structure of the world. But structure is not just a matter of piling levels of ever increasing time scales on top of each other. A key element of causal structure has to do with the *interactions* between regularities at different time scales. This interaction works in a bottom-up fashion such that, for example, the fast changing regularities governing contour, orientation, and so on help you become more confident that what you are looking at is really a nose belonging to an enduring face. It also works in a top-down fashion such that the longer-term regularities governing faces (e.g., they tend to be attached to headed bodies) assist in recovering fast scale changes in the input from the face (for example, the shadow cast by the nose as the body moves). For there to be such interactions, there must be extensive passing of messages up and down the different levels of the hierarchy. Understanding these messages is crucial to understanding how perceptual inference works (Lee and Mumford 2003).

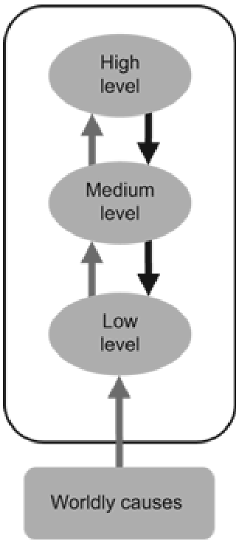
Using a development of the Bayesian story, Friston and his colleagues (Friston and Kiebel 2009) provide a computational model that exemplifies message passing across levels. A bird listening to another bird's song is extracting the fast time scale modulations of the song and can use that over time to extract slower time scale regularities about the size and strength of the other bird; perhaps stronger birds sing more distinctly and more forcefully for longer. But conversely, if an assumption is made about the size and strength of the singing bird, then that will help extract nuances in the fast scale dynamics of the song, which could otherwise be lost in noise. Low-level, fast scale regularities help choosing among hypotheses higher up and higher-level hypotheses about slower regularities work as control parameters on low-level regularities.

Top-down and bottom-up message passing of this type tie the levels of the perceptual hierarchy together. It is not as if the phenomenology of visual or auditory perception itself is just as of a causally shallow sensation that we can then subsequently label with categories for progressively deeper causal structures—

this is the kind of picture one would expect if processing at each level was in some sense complete and message passing was only a matter of sending a fully processed product to the next level for categorization. The picture is instead much more interactive with strong top-down modulation of lower level activity. That is, variant perception itself is steeped in causal structure. We find it hard to completely divorce the perception of the changing light and shadows from perception of which object it is, and the message passing throughout the perceptual hierarchy reflects this.

This picture of how the levels of the perceptual hierarchy connect thus depends on extensive message passing between levels. There are top-down expectations given slower time scale regularities and faster p. 32 time scale processing ↳ somehow sends messages in a bottom-up fashion that can guide higher-level processes. Figure 5 provides a schema of this initial idea; in the next chapter, more sophisticated versions will explicate what happens in the meeting between input and expectations, where prior expectations come from, and what happens next.

**Figure 5.**




The perceptual hierarchy, first version. Processing of causal regularities at different time scales influence each other in a bottom-up–top-down fashion. Sensory input (dark grey arrows pointing up) is met with prior expectations (black arrows pointing down) and perceptual inference is determined in multiple layers of the hierarchy simultaneously, building up a structured representation of the world. This figure simplifies greatly as there are of course not just three well-defined hierarchical levels in the brain. A later version of the figure will nuance the description of message passing between levels; in particular, the dark grey arrows will be re-labelled as “prediction error”.

This busy pattern of concurrent message passing is central to the hypothesis testing mechanism that will emerge in the next chapters, and only then can the force of the perceptual hierarchy really be appreciated.

## Additional constraints on hierarchical inference

In the light of the notion of the perceptual hierarchy, we can now revisit the issue concerning the need for p. 33 additional constraints on perceptual inference. ↳ In order to prioritize between different hypotheses about the causes of sensory input, the system needs to appeal to prior belief. But prior belief needs to be better than raw guessing, and our account of it will be circular if the story reduces to prior belief being directly based on the very thing we are trying to understand, namely perceptual inference. The problem was, that is, to account for prior belief without circularity. With the reciprocal message passing in the perceptual hierarchy we can do something to situate prior belief, without yet solving the problem.

Some prior belief is embodied in the expectations passed down from higher levels. To use the birdsong example again, if the bird expects the birdsong to come from a strong singer then inference for the extraction of individual notes at faster time scales lower down in the hierarchy can be guided along by those longer term expectancies. The prior expectations are therefore pulled down from what has previously been learned best at higher levels of the hierarchy (this is called empirical Bayes; for a brief introduction, see Bishop 2007: Ch. 3.5). This can happen in a cascading manner where very high-level expectations help shape many levels below but filtered through intermediate levels. This means the required additional constraint is not extracted directly from the sensory signal, which would lead to the threatening circularity or hopeless attempts at bootstrapping the process.

Helmholtz mentions an interesting case of a prior that can serve as an illustration, in this case of learned long-term term visual expectations of depth and colour in different parts of our visual field. He observes that the clouds in the sky have less depth perspective than objects on the ground, and that the colours of the objects on the ground appear to change depending on whether they are close by or far away. He rather charmingly seems to have tested this by sticking his heads between his legs:

> It appears that, when the head is turned upside down, the clouds get real depth while objects on the ground appear more like a painting on a vertical surface, much like the clouds in the sky [normally look]. In this case, colours also lose their relation to near and far objects and appear to us with their original differences. (Helmholtz 1867: 432)

That is, when the clouds appear in the lower visual field they immediately gain more perceived depth, and when objects normally on the ground appear in the upper visual field they lose depth and also the modulation of colour that is determined by depth cues. Many will have experienced this kind of effect when flying above the clouds and noting the unusual depth and beauty they seem to acquire when viewed from this perspective. The long term expectation for depth in the lower part of the visual field allows extraction of information from objects placed there, and the lack of this expectation for the upper half restricts our ability to extract depth information, even for familiar objects placed there.

p. 34 This idea of hierarchical, nested inference provides the first step for explaining the needed additional constraints on perceptual inference. It is a version of a standard move in discussions of Bayes when the question arises "where do the priors come from?" (Kersten, Mamassian et al. 2004; Friston 2005). It is not very satisfactory if priors are set entirely subjectively, and if not set entirely subjectively, then it seems we must go beyond the Bayesian framework to provide them. With the hierarchy, and the notion of empirical Bayes, we can just say they are extracted from higher levels.

But obviously, this can only be the first step of the explanation. The second step has to concern how these top-down priors are arrived at and how they are shaped over time. That step of the explanation has to show that the prior knowledge embedded in higher hierarchical levels is not given over to raw guessing. The explanation, as we will see in the next chapter, is that the priors are themselves guided by a particular kind of feedback signal stemming from processing of the incoming sensory signal.

A neat explanatory circle then seems to transpire: top-down priors guide perceptual inference, and perceptual inference shapes the priors. Couched in such simple terms, this circle is clearly not fit for underpinning perceptual inference: messages can be passed around in a completely idle fashion, never leading to perception of the world. The trick is to conceive of the two steps as being performed consecutively (it is not a circle as much as a spiral, as it were), and to give a particular, predictive flavour to the whole story. This is the job of the key prediction error minimization mechanism, which I will discuss in the next chapter and which will explain from where the priors come.

## On Bayes' rule

Bayes' rule is a simple result in probability theory and at the same time an extraordinarily powerful idea. It may be helpful to see first how this simple result comes about, and then see why it is co-opted as a paragon of rationality and scientific inquiry. This section then sets up some very minimal formal notation, which is used in later chapters.

Say we are interested in the joint probability of two random variables *D* and *C* taking the specific values *d* and *c*. As an example, say we are interested in how probable it is that the drought will break this year *and* that my computer breaks down this year. We can write this joint probability as

$$P(d,\ c)$$

One way of transforming this expression is like this:

$$P(d,\ c)\ =\ P\,(d|c)P\,(c)$$

p. 35 This makes intuitive sense because it just says that the probability of both events happening is the same as the probability that one event happens given the other one happens, times the probability that that other event happens. The probability that the drought breaks and my computer breaks can be found if we first find out how probable it is that the drought breaks given that we're in a year where my computer breaks, and then adjust that probability with the probability that my computer breaks in the first place. It is a way of chaining up the probabilities.

The order in which we chain up probabilities in this kind of expression does not matter. That is, we could just as well ask how probable it is that my computer breaks given we are in a year where the drought breaks, and then adjust that conditional probability with how probable it is the drought breaks in the first place. Thus we could say

$$P(d,c) = P\,(c|d)P\,(d)$$

But now we can combine the two right sides of these equations, because they are both equal to *P*(*d, c*), thus

$$P\,(d|c)P\,(c) = P\,(c|d)P(d).$$

It is simple then to divide with *P*(*c*) on both sides

$$P\,(d|c)P\,(c)/P\,(c) = P\,(c|d)P\,(d)/P\,(c),$$

and then cleaning up the left side

$$P\,(d|c) = P\,(c|d)P\,(d)/P\,(c),$$

which is none other than Bayes' rule itself. For our example, it says that the probability that the drought breaks given the computer breaks is equal to the probability that the computer breaks given the drought breaks, multiplied by the probability that the drought breaks in the first place, and then divided by the probability of the computer breaking in the first place.

This rule could well have lingered as just a result of probability theory but Bayes, Laplace, and others focused on it in attempts to think about how one should update belief in the light of new evidence. The next step is then to understand why this simple result should be seized upon for these purposes. We will distinguish between a *model* and a *hypothesis* in the sense that a model can entertain a number of competing or alternative hypotheses. For example, I could model outcomes in terms of the toss of a coin or the throw of a die. The coin and die correspond to models of observed outcomes. For each model there are a number of hypotheses—for example, the best hypothesis for explaining my observations under a coin model is that the
p. 36 coin came ↳ up tails. Throughout this book, we will be dealing largely with hypotheses, $h$, under a model, $m$, that is assumed by the agent in question. For present purposes, we set aside models and focus on hypotheses.

So, for application of Bayes' rule, we want to consider two things, a hypothesis, $h$, and some evidence, $e$. How strongly is the hypothesis supported by the evidence? Intuitively, that depends on two things. First, how tightly the evidence fits with the hypothesis, and second, how probable the hypothesis is in the first place. These two elements mirror the way we think critically about such matters, even if we often get the actual computations wrong when trying to do it.

Say we are confronted with a conspiracy theory about the 9/11 World Trade Centre attacks. The hypothesis of a massive, covert, state-driven conspiracy explains the evidence incredibly well: it is highly likely that we would have observed the bombings and much of the other related evidence, *given* there really was such a conspiracy. That is, the probability of the evidence is high, conditional on the conspiracy hypothesis, $P(e|h)$. This is presumably why some people begin to consider this kind of conspiracy theory.

But then we quickly think of how probable it is that there would be such a conspiracy in the first place, without considering this particular evidence. Of course, this probability, $P(h)$, is absolutely minuscule. So we say that even though the conspiracy hypothesis would certainly explain a lot, including snippets of evidence that competing hypotheses cannot explain, we should not believe it because it is just so very unlikely in the first place. But this is just a way of running through Bayes' rule: we multiply the (high) *likelihood* of the evidence given the hypothesis by the (minuscule) *prior probability*. We are interested in whether we should believe the conspiracy theory, $h$, given the evidence at hand, $e$ (*i.e.*, the attacks and so forth), that is, we ask "$P(h|e)$?", and we answer "$P(e|h)P(h)$".

Bayes' rule thus captures the two key elements we employ when we adjust our belief in a hypothesis given some new evidence. It may be instructive to run through for yourself an example where the likelihood $P(e|h)$ is very low, but where the prior probability is very high, as well as an example where there is both high likelihood and high prior.

The next step is then to compare the resulting *posterior probability,* $P(h|e)$, with the posteriors for other hypotheses, $h'$. Then we can rank hypotheses and end up believing, and acting upon, the one with the highest posterior probability.

It is fortunate that this natural way of reasoning about belief is captured by a result from probability theory because it tells us that doing so is in some way rational (for example, in a betting scenario, relying on probability theory ensures one doesn't irrationally accept certain bets guaranteed to lose one money).

Notice that the denominator ($P(e)$) of Bayes' rule is ignored in this heuristic treatment. This is because it is
p. 37 normally written in a slightly different way, ↳ such that its role is to normalize the resulting posterior

probability to a value between 0 and 1, as required in probability theory (this is why probabilities are normally reported as for example, 0.1 or 0.8 but never as 7 or 42). This technical detail can be brought out by noticing that $P(e)$ is the *marginal probability*, that is, the sum of the probabilities of $e$ conditional on all hypotheses.

For the purposes of perception, and the main themes of this book, the key idea is that $e$ is the sensory input and $h$ and $h'$ are competing hypotheses, maintained in the brain, about the state of the world. The hypothesis that is selected determines perceptual content, and therefore belief and action are determined by how the brain assesses likelihoods and weighs that by priors; the next chapter looks at the mechanism for achieving this.

## Summary: hierarchical neuronal inferential mechanisms

This chapter presented the problem of perception in terms of causal inference, and presented a Bayesian approach to this problem. The idea that perception is inferential may seem rather intellectualist but I motivated the need for inference by noting the phenomenon of binocular rivalry, which seems impossible to explain without use of inferential mechanisms. I also briefly indicated why I think it is appropriate to describe the brain as engaged in inference-making.

I then argued that the Bayesian notion of perceptual inference has the resources to capture not just our ability to use perception to recognize and categorize states of affairs in the world but also to capture the phenomenological richness of perceptual experience. The central tool for this is the notion of the perceptual hierarchy.

The reciprocal, top-down–bottom-up message passing in this hierarchy seems able to accommodate the variant and invariant aspects of perception as well as the first-person perspectival nature of our perceptual experience. Finally, the perceptual hierarchy allows us to situate the prior beliefs that perceptual inference must avail itself of in order to guide its selection of the best hypothesis. This hypothesis corresponds to the parameters of a probabilistic model encoding beliefs about hidden states in the world. This model has a hierarchical form such that its parameters are selected to provide an explanation for the world, at multiple levels of description. We noticed also the distinction between selecting the parameters of a model (as in the perceiving brain that selects a hypothesis) and actually selecting a model that could contain different parameters (selecting a brain). The latter may be a long-term process at a developmental or even
p. 38 evolutionary timescale. In this book, ↳ we will focus primarily on selecting the parameters of a particular model (brain), where each setting of the parameters represents a distinct hypothesis.

The perceptual hierarchy will play a central role in many of the discussions in the following chapters. It is built up out of replications of the prediction error minimization mechanism we will see in the next chapter, and the message passing is a crucial part of how this mechanism works.

This chapter has prepared the ground for the main message of this book. Once we see that perception is inferential and happens in a causal hierarchy we will be able to see how perception is a matter of prediction error minimization. The task that lies ahead in the next two chapters is then to explore how the neural machinery in the brain can realize probabilistic, causal inference. An important step is to reconceive this kind of inference such that it becomes more obvious that an organ like the brain can be an *inferential mechanism*, and how this speaks to the full richness of perceptual experience.

# Notes

*Page 14.* [“One possibility is that the additional…”] By referring to these mere regularities I have in mind the kind of law-like statements proposed as a model of non-inferential knowledge by Armstrong and discussed as candidates for a theory of content by Fodor (Armstrong 1973; Fodor 1990).

*Page 14.* [“But even if at some level of…”] The idea that there is a normative aspect to the problem of perception is put most forcefully by (Kripke 1982) in his discussion of linguistic meaning; Chapter 8 will have more discussion of this work.

*Page 15.* [“Consider this very simple scenario…”] The simple scenario of someone trying to make sense of the external world from inside a room by extracting statistical regularities might be called the “Reichenbach Room” after Hans Reichenbach’s example of the “cubical universe” in his *Experience and Prediction* (1938). Eliott Sober (2011) discusses this case and anticipates elements of the notion of active inference, discussed in Chapter 4 below.

*Page 18.* [“While the Bayesian, inferential approach…”] So far the account of Bayesian theories of perception is greatly simplified. Researchers have extensively worked out theoretical aspects of Bayesianism as applied to perception and cognition, and tested many of them in the lab (for reviews, see Clark and Yuille 1990; Knill and Richards 1996; Kersten, Mamassian et al. 2004; Chater, Tenenbaum et al. 2006). In addition to Bayesian theories that focus on perception, there are influential, more cognitive applications of Bayes to how we learn concepts, acquire language, and grasp causal relations, and more generally how the brain manages to generalize from sparse samples and apply that knowledge under uncertainty (Tenenbaum and Griffiths 2001; Tenenbaum, Kemp et al. 2011).

*Page 19.* [“Some centuries later…”] Helmholtz argues for the need for inference in perception on similar grounds as I do here. He p. 39 does not use rivalry as his initial ↳ example but instead the ambiguity of movement of light on the retina when we look to the side in the normal way vs. when we forcibly use the fingers to push the eyeball (Helmholtz 1867: 428). Helmholtz discusses rivalry in attentional but also rather inferentialist terms towards the end of this edition of *Physiological Optics*, beginning at page 766. Ibn al Haytham acknowledges the need for perceptual inference because he is aware that there are optical distortions and omissions of the image hitting the eye, which without inference would make perception of similarity and difference, colour, transparency, and written language impossible (Lindberg 1976; Hatfield 2002). Thus al-Haytham (ca. 1030; 1989) says “not everything perceived by the sense of sight is perceived by pure sensation; rather, many visible properties are perceived by judgement and inference [syllogism] in addition to sensing the visible object’s form, and not by pure sensation alone” (II.3.16); “the shape and size of a body…and such-like properties of visible objects are in most cases perceived extremely quickly, and because of this speed one is not aware of having perceived them by inference and judgement” (II.3.26).

*Page 20.* [“It is a surprising effect…”] Rivalry was as lively a topic of discussion around the mid-1800s as it is today. Wheatstone for example chastizes the experimental philosopher Thomas Reid for claiming he experiences something like a fused blend and not rivalry (Wheatstone 1838: §14). Helmholtz comments on the controversies and remarks on the significant individual differences in bistable perception (Helmholtz 1867: 437–8).

*Page 23.* [“As we will see later…”] Helmholtz beautifully anticipates the core Bayesian story about rivalry and bistable perception in general. He notes that sometimes “numerous comparisons and their corresponding interpretations are possible for the sensory impression. In such cases the explanation [of the impression] vacillates such that the observer has different experiences, one after another, for the unchanging retinal image” (Helmholtz 1867: 438; my translation).

*Page 23.* [“The proposal is that the brain…”] A good place to begin on the difference between personal level and subpersonal level explanation is (Davies 2000). There is a useful overview of functionalism in (Braddon-Mitchell and Jackson 2006) and a comprehensive study of the philosophy of science of neuronal mechanisms and levels of explanation (Craver 2007). Different interpretations of “realization” of a functional role by some physical property leads to different metaphysical conclusions (Melnyk 2003; Kim 2008).

*Page 24.* [“Very few people would claim…”] For further discussion of the worry about being too intellectualist and about neuroanthropomorphism, see (Chater, Tenenbaum et al. 2006); for criticism (Colombo and Seriés 2012); for a kindred defense see Rescorla (in press); see also (Phillips 2012) who responds to this kind of challenge in a discussion of Jaynes’ early approach to probability.

*Page 25.* [“In many ways, this…”] The relation between inference to the best explanation and Bayes is discussed in the second edition of Peter Lipton’s classic book (2004) on the topic.

*Page 26*. ["This is an important task…"] There is a deeper way to engage this kind of issue about the determination of
p. 40 perceptual content, which has to do with whether ↳ the perceptual relation determines perceptual content or whether it is a more austere relation to things in the world; see (Schellenberg 2010) whose views seem to me congenial to the view argued for here; see also discussion in Chapter 11.

*Page 28.* ["The brain responds to this importance of causal…"] For some evidence of the perceptual hierarchy in the auditory domain, see (Wacongne, Labyt et al. 2011) who showed that a low level absence of sensory input can be surprising and more surprising than a merely different low level input, a pattern which is best explained by reciprocal message passing between low level input and expectations anchored in higher level representations of slower regularities. Evidence is also given by Harrison et al. (Harrison, Bestmann et al. 2011), who finds "that visual and parietal responses are released from the burden of the past, enabling an agile response to fluctuations in events as they unfold. In contrast, frontal regions are more concerned with average trends over longer time scales within which local variations are embedded. Specifically, [there is] evidence for a temporal gradient for representing context within the prefrontal cortex and possibly beyond to include primary sensory and association areas." For the temporal structure of the prefrontal cortex, see also (Foster 2001).

*Page 28.* ["Perceptual inference happens in this…"] In this section I describe aspects of what I call the perceptual hierarchy. In addition to the work by Friston and colleagues that I focus on there are important developments of Hierarchical Bayesian Models (HBM), which speak to many of the issues and examples I have mentioned here (Tenenbaum, Kemp et al. 2011).

*Page 30*. ["It is thus possible to think of…"] For the notion of first-person perspective and perceptual content see (Metzinger 2009) as well as the discussion in (Schellenberg 2008; Jagnow 2012).

*Page 30.* ["The causal hierarchy is thus…"] Here I propose that our tendency to reality test is anchored in the different levels of the perceptual hierarchy. More abstractly, this seems to relate to our sense that there is a difference between appearance and reality, between how things seem and how they are. Metzinger describes this in terms of the emergence of a degree of opacity in our representational content and suggests in a way congenial to my treatment that hereby "[t]he difference between appearance and reality itself becomes an element of reality, and it can now be acted upon or thought about, attended to, and made the object of closer inspection." (Metzinger 2004: 176).

*Page 33*. ["Helmholtz mentions an interesting…"] My translation of Helmholtz, the original quote is "Ja es kommt wohl vor, dass bei umgekehrtem Kopfe die Wolken richtige Perspective bekommen, während die Objecte der Erde als ein Gemälde auf senkrechter Fläche erscheinen, wie sonst die Wolken am Himmel. Damit verlieren auch die Farben ihre Beziehung zu nahen oder fernen Objecten, und treten uns nun rein in ihren eigenthümlichen Unterschieden entgegen."

*Page 34*. ["Bayes' rule is a simple result…"] There is a mountain of literature and research on Bayes' rule. It all begins with this simple intuition, that it captures something basic about weighing evidence and adjusting belief (two good places to start are Howson and Urbach 1993; Bovens and Hartmann 2003). There is also much research on how poor we are at conscious Bayesian reasoning (Kahneman, Slovic, et al. 1982; Gigerenzer and Selten 2001).