Imaginary Foundations*

Wolfgang Schwarz
Draft, 24 June 2015

Abstract. Our senses provide us with information about the world, but what exactly do they tell us? I argue that in order to optimally respond to sensory stimulations, an agent's doxastic space should have an extra, "imaginary" dimension of possibility; perceptual experiences confer certainty on propositions in this dimension. To some extent, this vindicates the old-fashioned empiricist idea that all empirical knowledge is based on a solid foundation of sense-datum propositions, but it avoids most of the problems traditionally associated with that idea. The proposed model also explains why experiences appear to have a non-physical phenomenal character, even if the world is entirely physical.

1 Learning from experience

Through the window I can see that it is still raining. A stream of water is running down the street into the gutters. But can I tell, just be looking, that it is water? Couldn't it be a stream of vodka? To be sure, that is an outlandish possibility. But if for whatever reason I had taken the vodka hypothesis seriously before looking outside, my visual experience wouldn't put me in a position to rule it out. So if we define the information provided by my visual experience in terms of the possibilities the experience allows me to rule out, then the information I receive from my senses does not entail that there is a stream of water on the road. Nor does it entail that it is raining. What looks like rain could be a setup for a movie scene. My windows could have been replaced with sophisticated LCD screens. Again, my visual experience by itself does not put me in a position to rule out these possibilities.

This line of thought naturally leads to the old empiricist idea that the information we receive from our senses is in the first place not information about the external world, but about a special, luminous, internal realm of appearances or sense data: the possibilities I can rule out are all and only the possibilities in which things do not appear as they actually do. Yet this view also faces serious problems. Aren't we often ignorant or

* Ancestors of this paper were presented at the University of Bielefeld and the ANU in 2006, the University of Cologne in 2010, the University of Melbourne in 2013, and the University of Leeds in 2014. I thank the audiences at these events for helpful feedback. Thanks also to David Chalmers, Colin Klein and especially Christian Nimtz and Daniel Stoljar for detailed comments on earlier versions.

mistaken about how things appear? How could everything we know about the world be inferred from facts about appearance? How are appearance facts supposed to fit into a naturalistic account of the mind?

So perhaps we started with a bad question. Let's drop the assumption that perceptual experiences put us in a position to exclude possibilities, conclusively and independently of our background beliefs. My experience, perhaps together with my background beliefs, merely allows me to conclude tentatively and defeasibly that it is raining. In general, experiences combine with background beliefs to confer degrees of plausibility or probability to various claims about the world.

But things are not so easy. To bring out why, let's try to model the present proposal in the framework of Bayesian epistemology. Here we assume that beliefs come in degrees, measured by numbers between 0 and 1. We further assume that rational degrees of belief satisfy the mathematical conditions on a probability function. How should these probabilities change under the impact of perceptual experience? Classical Bayesian epistemology suggests the following answer. For each type of perceptual experience there is a proposition E such that whenever a rational agent has the experience, then her new probability should equal her previous probability conditional on E; i.e. for all A, $P_{new}(A) = P_{old}(A/E) = P_{old}(A \wedge E)/P_{old}(E)$, provided $P_1(E) > 0$. In this case P_{new} is said to come from P_{old} by *conditionalising on E*. Since $P_{old}(E/E) = 1$, it follows that the new probability of E is 1. So E can hardly be an ordinary proposition about the world. When I looked through the window, I didn't become absolutely certain that it is raining water. (Indeed, my experience *raised* the probability of the hypotheses that it is raining vodka; the new probability is negligible only because it was negligible in the outset.) Again, we seem forced to postulate a mysterious realm of sense-datum propositions to serve as objects of conditionalisation.

To avoid this commitment, Richard Jeffrey proposed an alternative to the classical Bayesian picture, *radical probabilism* ([Jeffrey 1965: ch.11], [Jeffrey 1992]). Radical probabilism rejects the idea that subjective probabilities require a bedrock of certainty. To use a well-known example from [Jeffrey 1965], imagine you catch a glimpse of a tablecloth in a poorly lit room. According to Jeffrey, the direct effect of this experience on your beliefs may be that you come to assign credence 0.6 to the hypothesis that the cloth is green and 0.4 to the hypothesis that it is blue; these probabilistic judgments need not be inferred from anything that became certain.

In general, Jeffrey assumes that an experience is directly relevant to some propositions and not to others. Suppose $E_1, \ldots, E_n$ is a list of pairwise exclusive and jointly exhaustive propositions whose probabilities change in response to an experience so that their new values are $x_1, \ldots, x_n$ respectively. If the experience is directly relevant only to $E_1, \ldots, E_n$, then probabilities conditional on these propositions should be preserved: for all A and E_i, $P_{new}(A/E_i) = P_{old}(A/E_i)$. It then follows that the new probability of any proposition A

is given by

$$P_{new}(A) = \sum_i P_{old}(A/E_i) \cdot x_i.$$

This transformation from P_{old} to P_{new} is known as *Jeffrey conditionalisation* or *probability kinematics* or *generalised conditionalisation* – but as we will see in a moment, that last name is rather misleading.[1]

At first glance, Jeffrey's model seems to deliver just what we were looking for. Instead of assuming that for each perceptual experience there is a sense-datum proposition E rendered certain by the experience, we only need to assume that there is a distribution of probabilities $x_1, \ldots, x_n$ over some partition $E_1, \ldots, E_n$ of ordinary propositions such that when a rational agent has the experience, then her degrees of belief evolve by the corresponding instance of Jeffrey conditionalisation.

More concretely, we could assume that every experience is associated with a proposition E that captures how the experience intuitively represents the world as being, and a number x giving the degree of belief the agent should assign to this proposition; $E_1, \ldots, E_n$ and $x_1, \ldots, x_n$ in Jeffrey's rule then reduce to $E, \neg E$ and $x, 1-x$. We might still face a challenge to explain why this response is justified: why it is OK to tentatively assume that the world is as it appears to be. But at least we seem to have a *structurally* sound model of belief change that frees us from the implausible commitments of the classical model.

Unfortunately, the present model won't do either, since the new probabilities $x_1, \ldots, x_n$ for $E_1, \ldots, E_n$ should depend on the previous probabilities. Your new beliefs about the colour of the tablecloth should be sensitive not only to your visual experience in the dimly lit room, but also to background information about (for example) the colour of other tablecloths in the house. So how does an experience together with an agent's previous belief state determine the "inputs" to a Jeffrey update: the evidence partition $E_1, \ldots, E_n$ and the associated probabilities $x_1, \ldots, x_n$?

This is sometimes called the *input problem* for Jeffrey conditionalisation. The problem was already raised by Carnap in his 1957 correspondence with Jeffrey (published in [Jeffrey 1975]). Carnap, an early advocate of classical Bayesianism, reports that he had himself attempted to relax the Bayesian account along Jeffrey's lines but had given up because he couldn't find an answer to the input problem. Since then, nobody else has found a plausible answer either. It is widely thought that the problem simply can't be solved.[2]

1 For ease of exposition, I have assumed that the experience is directly relevant only to a finite partition $E_1, \ldots, E_n$, but the formula is easily extended to infinite cases; see [Diaconis and Zabell 1982: sec. 6].

2 As Skyrms [1980] points out, the effect of Jeffrey conditionalisation can be mimicked by standard conditionalization on probabilistic propositions: what you learn for certain is that you are 60 percent certain that the tablecloth is green. This account does not require sense-datum propositions, but it inherits the input problem for Jeffrey conditionalisation: how does your perceptual experience

To get a sense of the difficulties, consider a version of the tablecloth scenario in which you look *twice* at the cloth in the dimly lit room, from the same point of view. Suppose your first experience increases your credence in the hypothesis that the cloth is green from 0.3 to 0.6. Absent unusual background beliefs, your second experience should not significantly alter your beliefs about the cloth's colour. Intuitively, this is because the second experience is in all relevant respects just like the first and thus provides little new information. (By contrast, if you had had two equally inconclusive but very different experiences of the cloth, the second would have carried more weight.) The problem is that this fact about the two experiences may not be recoverable from your credence prior to the second experience together with a specification of the new experience. To be sure, if every experience had a "phenomenal signature" that (a) distinguished it from all other experiences and (b) was infallibly revealed to everyone who has the experience, then we could consult your credence function to see if you recently had the same type of experience. But the whole point of radical probabilism was that we wanted to do without such phenomenal signatures.[3]

Jeffrey, in any case, never gave an answer to the input problem. His radical probabilism is silent on how perceptual experiences together with previous beliefs and possibly other factors yield new probabilities $x_1, \ldots, x_n$ for a partition $E_1, \ldots, E_n$ such that probabilities conditional on each E_i are preserved. All it says is that *if* somehow or other this happens, *then* the new probabilities ought to result from the old ones by the relevant instance of Jeffrey conditioning. But that much is a simple consequence of the probability calculus. Jeffrey's account therefore doesn't provide a substitute for conditionalisation as the second norm of Bayesian epistemology. His alternative collapses into the first norm, probabilistic coherence.

This leaves a serious gap in Bayesian epistemology (as noted by Carnap and reiterated e.g. in [Field 1978] and [Christensen 1992]). The demands of epistemic rationality go well beyond probabilistic coherence. In particular, there are norms on how one's beliefs may

together with your prior belief state determine the evidence proposition?

3 The present point is inspired by [Garber 1980], where it is shown that the specific proposal in [Field 1978] fails for repeated experiences. [Hawthorne 2004] presents a model that gets around the problem by making the input parameters to Jeffrey conditionalisation depend not only on present experience and old probabilities, but also on earlier experiences; the dynamics of rational credence is thereby rendered non-Markovian. Other challenges to solving the input problem arise from the holistic character of evidential support; see [Christensen 1992], [Weisberg 2009], [Wagner 2013] and the discussion of parochialism in [Jeffrey 1988]. The basic worry here is that if probabilities are only defined over ordinary external-world propositions, then it may be impossible to find a non-trivial evidence partition $E_1, \ldots, E_n$ that screens off the experience from all other propositions in the sense that $P_{new}(A/E_i) = P_{old}(A/E_i)$. [Weisberg 2009] also points out that a result in [Wagner 2002] seems to entail that the failed proposal of [Field 1978] is the only systematic answer to the input problem that satisfies a desirable commutativity condition (roughly, that it makes no difference to the final probabilities which of two experiences arrives first).

change through perception. When chemists use litmus strips to test whether a solution is basic or acidic, they are not free to change their beliefs in any way they please in response to the outcome. My visual experience of the rain supports the hypothesis that it is raining, but not that it is snowing or that Tycho Brahe was poisoned by Johannes Kepler. ("How do you know?" – "I looked through the window".)

Even if there were no such norms, we would have a gap in Bayesian *psychology*. A psychological model of rational agents should have something to say on how belief states change under the impact of perceptual experience. If this could not be done within the Bayesian framework, we should conclude that something is wrong with the framework. But of course the problem isn't internal to Bayesianism. The general problem, illustrated by examples like the repeated tablecloth experience, is that if beliefs only pertain to ordinary external-world propositions, then an agent's rational belief state does not seem to be determined by her perceptual experiences together with her previous belief state. Something else plays a role. We need to know what it is and how it works.

2 The "Bayesian model" of perception

To make progress on the issues raised in the previous section, it may help to change perspectives and think about how we would design an ideal agent. Imagine we are to build a robot whose task is to find certain objects in its environment – mushrooms perhaps, or tennis balls, or landmines. To this end, the robot has a database in which it can store probabilistic information about the environment. It also has sense organs to receive new information. How should the probabilities in the database change in response to signals received from the sense organs?

A sense organ is a physical device whose internal state systematically and reliably depends on certain features of the environment. Let's assume our robot has a visual sense organ consisting of a two-dimensional array of photoreceptors, like in the human eye. When hit by electromagnetic waves of suitable wavelengths, each photoreceptor produces an electrical signal. Different colours, shapes and arrangements of objects in the environment give rise to different patterns of light waves activating the photoreceptors, which in turn lead to different total signals produced by the sense organ – i.e., to different patterns of electro-chemical activity in the "output" wires of the device. (This is all I mean here by a *signal*).

Ideally, we would be able to read off the exact colours, shapes and arrangement of objects in the environment from the signal produced by the sense organ. But this is not possible, because different configurations of the environment lead to the very same activation of photoreceptors and thus to the very same signal: from certain angles, a three-dimensional cube might create the very same signal as a two-dimensional hexagon; a small cube at close distance might cause the very same signal as a larger cube further away; a convex shape with light from above might create the same signal as a concave figure with light from below; a red cube under white light might create the same signal as a white cube under red light.

So the functional architecture of a sense organ only determines, for each signal S a range of alternative hypotheses about the environment $E_1, \ldots, E_n$ that could be responsible for S. Typically, some of these environmental conditions will be much more common than others. If our robot traverses the surface of the Earth, it will mostly find itself in situations where roughly white light is coming roughly from above. Nevertheless, the robot arguably shouldn't become certain that a particular one of $E_1, \ldots, E_n$ obtains, giving zero probability to all the others. A better idea is to implement a form of Jeffrey conditionalisation, where the new probabilities $x_1, \ldots, x_n$ over $E_1, \ldots, E_n$ reflect the ecological relative frequency (or, if you want, objective chance) with which the conditions obtain when the signal is produced.[4]

But that is still not an optimal solution. As mentioned in the previous section, the new probability assigned to the E_i's should be sensitive not only to the sensory signal but also to the old probabilities. For example, suppose signal S occurs just as often under condition E_1 as under E_2, so that the ecological frequencies x_1 and x_2 are the same. Suppose further that before the arrival of S the robot has received information that supports E_1 over E_2. On the present account, the new signal will wipe out this information, setting $P_{new}(E_1) = P_{new}(E_2)$. This is clearly not ideal. Relatedly, our robot should be able to learn whether it is in an environment where S generally goes with E_1 or E_2; in the present model, the x_i values are fixed once and for all.

These problems suggest a more versatile approach which puts the x_i values under the robot's epistemic control, so to speak. Let's give our robot subjective probabilities about what the environment might be like given that a particular signal occurs. That is, the robot's database stores conditional probabilities of the form $P(E_i/S)$ for various signals S and propositions E_i about the environment. Initially, before the robot has encountered relevant information, these conditional probabilities could match the objective frequencies; but now they can change. The robot can learn that in the presence of S, E_1 is more common than E_2, but also that under certain conditions the situation is reversed and S is more likely to go with E_2 than E_1. Moreover, receiving the signal S won't overwrite any previous evidence regarding $E_1, \ldots, E_n$. Formally, the update still takes the form of Jeffrey conditionalisation, where the $x_1, \ldots, x_n$ values are given by $P_{old}(E_1/S), \ldots, P_{old}(E_n/S)$.

4 This kind of architecture is advertised in [Skyrms 1984: 117f.]. As just introduced, the list of alternatives $E_1, \ldots, E_n$ may not be exhaustive, which is assumed in Jeffrey's rule; we could make it exhaustive by simply adding a catch-all element E_{n+1} = 'none of $E_1, \ldots, E_n$' with associated probability $x_{n+1} = 0$. Even better, we could take into account that sense organs can fail. $E_1, \ldots, E_n$ are the situations that would cause signal S *under perceptually normal conditions*, but we can't guarantee that S couldn't also be triggered "erroneously", say if the robot is hit with a bat. So we should set $x_{n+1} > 0$ and adjust the other probabilities accordingly.

But this instance of Jeffrey conditionalisation is equivalent to strict conditionalisation on S, so we may as well bypass the evidence partition $E_1, \ldots, E_n$ and simply say that for all A, $P_{new}(A) = P_{old}(A/S)$.

This looks promising. However, recall that a signal S is a type of electrochemical activation in the output wires of the robot's sense organs. Thus it is not clear what it even means to say that the robot assigns conditional probability x to some proposition A conditional on some signal S: the objects of subjective probability are usually understood to be propositions, not types of electrochemical activation.

So what we need is a mapping ρ from signals to propositions such that whenever a signal S arrives, the robot conditionalizes on the associated proposition $\rho(S)$. But what is $\rho(S)$? The most obvious candidate is the proposition that the robot presently receives signal S. On this model, when the robot receives a signal S from its sense organs, it becomes certain that it receives that very signal and adjusts its other beliefs in accordance with that information.

An odd consequence of this proposal is that our robot must now have sophisticated opinions about the electrochemical state of its own sensory organs. One might have thought that a robot in search of mushrooms wouldn't need to be trained in electrochemistry. Indeed, a little reflection makes clear that this is not actually required.

Suppose the robot's probabilities were originally defined only for certain propositions $\mathbf{R}$ about the macroscopic environment. The above considerations suggest that in order to determine how a sensory signal S should affect those probabilities, we need to expand the domain of the probability function by further elements $\mathbf{E}$ such that when a signal S arrives, the robot conditionalizes a corresponding proposition $\rho(S) \in \mathbf{E}$. For this to work, it does not really matter what, if anything, $\rho(S)$ says about the world, as long as the robot assigns suitable conditional probabilities to propositions $A \in \mathbf{R}$ given $\rho(S)$. In particular, it is not required that $\rho(S)$ correctly describes the electrochemical signal S.

The elements of $\mathbf{E}$ are "propositions" in the sense that they are bearers of subjective probability, but they do not need to capture genuine ways the world might have been. Suppose they don't. When the robot receives a signal S, what it learns about the world is then fully given by the change in its probabilities over $\mathbf{R}$.

Think of it this way. The robot's probability function is supposed to capture its probabilistic views about the world. The function should therefore assign probabilities to various ways the world could be. This is its primary purpose. But we also want these probabilities to change under the impact of sensory stimulations, and to this end it proves useful to add further elements to the domain of the probability function. The values assigned to these elements and (more importantly) to combinations of these elements with the original propositions are not meant to encode the robot's views about the world. Rather, they reflect the robot's dispositions to change its views about the world in response to perceptual signals. If the robot assigns greater probability to A & E_S

than to $\neg A \,\&\, E_S$, where E_S is the element $\rho(S)$ of $\mathbf{E}$ associated with signal S (i.e., the element whose probability is raised to 1 by signal S), then receiving S will make the robot assign greater probability to A than to $\neg A$. To play this role, the new elements don't need to distinguish genuine ways the world could be.

I do not deny that sensory signals carry information about the world. If the robot's sensory signal is triggered by an ordinary confrontation with a cube and causes the robot to assign high probability to the presence of a cube, it makes perfect sense to say that the signal represents the presence of a cube, or carries the information that there is a cube. Nonetheless, on the present model the robot does not conditionalize on this information. It does not become absolutely certain that it confronts a cube, nor does it become certain of any other proposition about the world.

Formally, the required extension of a probability function is a straightforward product construction. Take the simplest case where everything is finite. Let $\mathbf{R}$ be the set of real propositions we ultimately care about. Probability theory demands that $\mathbf{R}$ is closed under Boolean operations. It follows that we can identify the propositions in $\mathbf{R}$ with sets of "possible worlds": the atoms of the algebra. Now let $\mathbf{E}$ be an arbitrary set disjoint from $\mathbf{R}$ such that there is a one-one correspondence between $\mathbf{E}$ and the signals the agent can receive.[5] Each pair $\langle w, e\rangle$ of a possible world w and an element e of $\mathbf{E}$ is then an atom in the extended doxastic space; each set of such atoms is a bearer of probability.

Some terminological stipulations will be useful. I will call anything to which an agent's probability function assigns a value a *(complex) proposition*. Genuine ways things could be (members of $\mathbf{R}$) are *real propositions*; subsets of $\mathbf{E}$ are *imaginary propositions* – in analogy to complex, real, and imaginary numbers, and to highlight the fact that imaginary propositions are not ways the world could be and in that sense not genuine propositions at all. Individual members of $\mathbf{E}$ I will call *sense data*, since their role in the present model is similar to the role of sense data in the classical empiricist model (see section 4 below). In an agent's complex doxastic space, a real proposition A can be re-identified with the set of atoms $\langle w, e\rangle$ whose possible world coordinate w lies in A; similarly for imaginary propositions and sense data.

Complex propositions here behave a lot like *centred* propositions in the modelling of self-locating beliefs. Arguably, our doxastic space contains not only ordinary propositions about the universe as a whole, but also *self-locating propositions* concerning our own current place in the world. Even if you knew every truth about the world from a God's eye perspective, you might not yet know who in the world you are and what time is now. Thus the atoms in an agent's doxastic space are often modelled as pairs $\langle w, c\rangle$ of a possible world w and a centre c which may specify an individual and a time. In the

5 $\mathbf{E}$ is an "arbitrary" set because the identity of its members is irrelevant to the functional specification of the robot's cognitive architecture. In this respect, the framework of probability theory is a little artificial, since it forces us to make a choice.

resulting doxastic space, an ordinary proposition A about the world is then identified with the set of "centred worlds" $\langle w, c\rangle$ whose possible world coordinate w lies in A. (I will return to this analogy in section 5.)

Our model is not quite complete yet. To see why, imagine our robot is moving towards a cube. At time t_1 it receives a signal S which (by the robot's lights) indicates that the cube is about 5 metres away. At t_2, it receives another signal S' indicating that the cube is about 4 metres away. We don't want the robot to conclude that the cube is most likely *both 5 and 4 metres away*. Nor should it conclude that the previous signal was faulty. Rather, it should realise that the first signal indicated that the cube was 5 metres away *at the time, at* t_1, which is perfectly compatible with the distance now being 4 metres.

To model this sort of process, we first of all have to add time indices to our sense data: E_S-*at*-t_1, $E_{S'}$-*at*-t_2 (or E_S-*earlier*, $E_{S'}$-*now*), etc.[6] When signal S arrives at t_1, the robot conditionalises on the corresponding proposition E_S-*at*-t_1 (or E_S-*now*). As a result, $E_{S'}$-*at*-t_1 has probability zero, but $E_{S'}$-*at*-t_2 may still have positive probability. If the robot knows that it is moving towards the cube, then it might expect to receive signal S' at t_2 when it just received S at t_1. E_S-*at*-t_1 then raises the probability of $E_{S'}$-*at*-t_2. In addition to conditionalising on new sense data, a complete update step must also adjust the agent's internal clock. The details of this step are somewhat controversial; see [Schwarz 2012] for my own proposal and some alternatives. For the present topic, these details are not important. In fact, I will mostly ignore the time indices in what follows.

This completes my proposal. This is how I would design an ideal agent. The ideas I have outlined are indeed well-known in artificial intelligence (see e.g. [Russell and Norvig 2004], esp. ch.15), although computer scientists tend to make little fuss about the addition of imaginary propositions: they call them 'sense data', 'raw data', 'input strings' or the like, and rarely pause to discuss their representational features.[7]

Given the importance of perception to survival, one might expect that nature has implemented mechanisms not too far from the present ideal in ourselves and other higher animals. Research in cognitive science appears to bear out this expectation. Here the present model is known as *the Bayesian model of perception* and has proved successful in explaining and unifying many aspects of sensory processing in humans and other animals (see e.g. [Kersten et al. 2004] for a survey).[8]

6 More precisely, probabilities are now defined over the three-fold product $\mathbf{R} \times \mathbf{E} \times T$ (or a sub-algebra of that product), where $\mathbf{T}$ is a set of times.

7 There is also a superficial difference in presentation. Computer scientists usually take the agent's conditional probabilities $P_{old}(A/E_S)$ over real propositions A given sense data E_S as derived from inverse probabilities $P_{old}(E_S/A)$ of sense data given hypotheses about the world; the update then goes via Bayes' Theorem. Since the posterior $P_{new}(A)$ is often written simply as $P_{old}(A/E_S)$ (making the subscript 'old' pointless) the assumption that probabilities change by conditionalisation is implicit in the notation.

8 A caveat: in cognitive science, the postulated probability function that gets conditionalised on sense

3 Interpretation and implementation

I have outlined a model of how subjective probabilities change under the impact of sensory stimulation. The agent's doxastic space is extended by a new, "imaginary" dimension whose points are associated with sensory signals in such a way that when a signal S arrives, the agent assigns probability 1 to the corresponding imaginary proposition E_S, while the probability of real propositions is adjusted in accordance with their prior probability conditional on E_S. The model is in the first place a normative ideal: a design template for an ideal computational system. But there are also reasons to believe that it is an approximation to how perception works in humans and other animals. How does it bear on the questions raised in section 1?

The answer depends in part on how we interpret the probabilities that figure in the model. In philosophy and psychology, terms like 'belief' and 'credence' are often used in a demanding intellectualist sense tied to conscious thought and linguistic assertion. On this usage, an agent's credences are taken to measure something like their degree of conviction when contemplating various sentences, or their dispositions to bet on those sentences. The probabilities in our Bayesian model are not credences in this sense. The claim is not that whenever we open our eyes, we find ourselves convinced of some "observation sentence" which prompts us to form conscious new judgments about every other sentence in our language.

There is an alternative tradition, going back to Ramsey [1931], which essentially treats 'credence' or 'belief' as theoretical terms characterised by their role in a comprehensive psychological theory connecting an agent's perceptual inputs and behavioural outputs with functionally individuated internal states. Here, belief is not assumed to be a conscious (or potentially conscious) attitude towards a sentence in the agent's language. The agent doesn't need to have a language at all. The objects of credence are better construed not as representational vehicles but as possible ways things could be. To assign a subjective probability to such a proposition, you don't need to form an explicit judgment, nor do you have to "grasp" the proposition in any interesting sense. All that's required is that your cognitive state plays the right causal-functional role (see e.g. [Stalnaker 1984]).[9]

Now suppose our model of perception looks something like the Bayesian model outlined in the previous section. The model postulates a probability measure over a space of "complex propositions", but it does not postulate that these probabilities are the agent's

data is usually not identified with the subject's credence function. I think the "Bayesian model of perception" is *compatible* with this interpretation, but for reasons discussed in the next section the issue is not completely straightforward.

9 One might still speculate that in order for a state to play that role, it should involve some kind of syntactic, sentence-like encoding of the relevant proposition. Empirical evidence suggests that this is not true for the probabilities involved in perception; see [Knill and Pouget 2004].

degrees of belief. One might reasonably insist that the labels 'degree of belief' and 'credence' should be reserved for the agent's attitudes towards real propositions. From this point of view, the Bayesian model does not say that in response to perceptual experiences, the agent's credences change by conditionalising on sense data. What gets conditionalised on sense data is not the probability function that captures the agent's views about the world but an extension of that function that also captures the agent's dispositions to change her beliefs through perception. Since the dynamics of the extended probability function determines the dynamics of the fragment that deals with real propositions, the Bayesian model indirectly specifies how credences in real propositions change through perception.

Formally, this update will take the form of Jeffrey conditionalisation. That's because we can always find some partition of real propositions such that the new credence function results by Jeffrey's formula from the new probabilities over this partition.[10] More interestingly, we can find a non-trivial partition whenever there are questions on which the received signal sheds no light. In simple cases, the partition could even be fixed by the alternative environmental conditions $E_1, \ldots, E_n$ that could have produced the received signal S, with the weights $x_1, \ldots, x_n$ set by the corresponding extended probabilities.

Thus we have found an answer to the input problem for Jeffrey conditionalisation: we can explain how a given experience together with an agent's prior doxastic state determines the input parameters to a Jeffrey update – provided that the "prior doxastic state" includes the agent's extended probability function. If we only look at the agent's prior probabilities over real propositions, the problem can't be solved because there is no fixed way in which sensory signals affect propositions about the world.

As I said, it makes sense to reserve 'credence' and 'degree of belief' for attitudes towards real propositions. But this terminological decision does not mean that an agent's cognitive architecture must draw a sharp distinction between the two kinds of propositions and their associated probabilities. Computationally, all propositions are on a par. Our robot does not need a special database for probabilities over real propositions if it already has a database for the complex probabilities. From the robot's perspective, reality might simply have an extra dimension – an extra respect of similarity and difference. If we just look at an agent's attitudes towards real propositions, we therefore not only ignore "lower-level" features of the agent's cognitive state that capture how she would respond to sensory stimulations. We might also miss an aspect of the agent's own epistemic perspective.

Similar remarks apply to the intellectualist use of 'belief' and 'credence' in philosophy. At some point, a psychological model of sophisticated agents like us should have something

10 The partition may not be finite, so we might have to use the more general form of Jeffrey's formula mentioned in footnote 1.

to say about how we use words and sentences to systematically communicate and reason about ways things could be. If a sentence in an agent's language expresses a proposition p, and the agent's probability for p is low, we can expect that the agent will (ceteris paribus) have corresponding attitudes towards the sentence: she won't be inclined to assert it, she will feel little conviction when contemplating it, and so on. The agent might also have words or "concepts" to reason about complex and imaginary propositions. However, her language may not be rich enough to express every proposition to which she assigns a probability. Moreover, even if the agent has a sentence for every proposition in her doxastic space, these linguistic vehicles need play no direct role in the process by which her probabilities change in response to sensory signals. Focusing on intellectualist degrees of belief will therefore often miss important aspects of an agent's epistemic state and its dynamics.

The Bayesian model of perception is obviously only a small part of a comprehensive psychological or epistemological theory. It accounts for the distinctive effect perceptual experiences have on an agent's subjective probabilities on a very abstract, structural level. It does not say which real propositions are normally supported by which perceptual experience. It does not say whether those real propositions are qualitative or singular. More importantly, it does not imply that perceptual experiences are the only force that acts on subjective probabilities. For example, a different kind of update is plausibly called for when we act or deliberate about future actions (see [Skyrms 1990]). If the psychological theory is to be at all realistic, it should also have something to say about forgetting: unlike ideal Bayesian learners, we don't always remember everything we ever learned.

How much realism should we demand? As in many parts of science, it can be useful to work out a simple, highly idealised model, ignoring friction and air resistance. Such a model won't fit the real world all that well, but it can still capture interesting and robust patterns in the phenomena; in the present case, the model might further serve as a normative ideal. On the other hand, we should also investigate how reality deviates from the ideal. Here too one can often find interesting patterns.

If we tried to actually build our robot, with its central database of probabilities updated by conditionalising on sense data, we would quickly hit insurmountable problems. Conditionalising a high-dimensional probability function is a non-trivial, often intractable computational task. Several tricks have been suggested to simplify the task. For one thing, we could relax the need to store and compute precise probabilities. It might be enough to encode various *constraints* on probabilities: that p is more probable than q, that r is probabilistically independent from s, and so on (see [Jeffrey 1984]). In addition, we could compress and distribute the central database. Thus we could delegate the processing of input signals to more or less encapsulated sensory modules whose task is to develop a coarse-grained probabilistic model of the immediate present environment. The output of these modules could then be used to compute the system's overall probability for a given proposition, when the demand arises.

More radically, we could have sensory modules that, while internally employing Bayesian techniques, produce as output a single, concrete hypothesis about the environment. This output would then provide the basis on which the agent adjusts her probabilistic representation of the world; it would serve as the "signal" in the agent's update procedure. Computationally, it is much easier to find a single plausible interpretation of an incoming signal – a single guess about the environment – than to calculate to what extent the signal supports each such hypothesis over the others. Given the huge amount of data our senses constantly receive, it might well make sense for our sensory modules to focus on this simpler, non-probabilistic goal.

In general, it may be worth emphasizing that in the update model I have put forward, the "signals" need not be thought of as excitation patterns in receptor cells. We can allow for various types of pre-processing. At the very least, it makes sense to transform the original excitation patterns into a more efficient format (perhaps by using techniques from "predictive coding"; see [Huang and Rao 2011]). The (imaginary) *content* of a signal may not be more than a bare tag, but we should certainly want this content to be encoded in a sensibly structured, quasi-compositional way.

I just mentioned that it is a lot easier to compute a single plausible interpretation of a signal than to compute its precise probabilistic impact. It is also a lot easier to *act* on the basis of a single hypothesis about the environment than on a probability measure: calculating expected utilities is generally just as intractable as conditioning. For agents like us and our ancestors for whom it is often vital to respond very quickly to incoming signals, it can thus be useful to have a fast and simple "dogmatic" mechanism alongside the slower probabilistic architecture. If it looks like a rock is about to hit your head, your first priority is to duck. Later you can figure out whether the object is really a rock, where it came from, and what it reveals about the geology of your environment.

This might be part of the explanation why perceptual experiences often seem to present the world as being a particular way. When I look at the Müller-Lyer illusion, there is a sense in which my visual experience suggests to me that one line is longer than the other. This is a kind of "perceptual content" (indeed, it is was philosophers mostly have in mind when they talk about perceptual content), but it is clearly not what I conditionalise on, since I do not become certain that one line is longer. I know that the lines are the same length – but the fast-track mechanism that produces the dogmatic interpretation is not sensitive to this knowledge.

It is not my aim in this paper to speculate about how our nervous system approximates the Bayesian ideal. This is a task for cognitive science. The present remarks are only meant to illustrate what a more realistic model that takes into account our cognitive limitations might roughly look like, and how the required compromises might account for salient features of our psychology that are not predicted by the simple Bayesian ideal.[11]

4 Hardcore empiricism?

According to traditional, empiricist foundationalism ('hardcore empiricism', as Jeffrey calls it), our knowledge of the world rests on a foundation of infallible and indubitable beliefs about present experience. The rest of our empirical knowledge is constructed on this basis, using a priori connections between the verdicts of experience and ordinary claims about the world. If these connections are logical, we get the striking phenomenalist view that the world (or at least all we can ever know about it) is a logical construction out of sense data.

In the Bayesian model I have outlined, experiences also confer absolute certainty on a special class of propositions. But these imaginary propositions don't distinguish genuine ways the world could be. That's why the world can never prove them false. This also means that there is no intrinsic logical or semantic connection between imaginary propositions and real propositions. (For example, imaginary propositions don't say that *it appears that* p, from which one might tentatively infer that p.) To get from sense data to claims about the world, we need external bridge principles, encoded in the agent's prior probabilities. If you receive a signal S and thus conditionalise on a corresponding imaginary proposition E_S, your new credence in a hypothesis H will be high if and only if your prior conditional credence for H given E_S was high.

In the terminology of Bayesian confirmation theory, we might say that E_s *subjectively confirms* or *supports* H for an agent with credence function P_{old} if $P_{old}(H/E_s) > P_{old}(H/\neg E_s)$. However, the mere fact that a belief is supported by sensory evidence relative to the agent's earlier beliefs does not automatically make the belief reasonable or justified. For one thing, the belief should in some sense be "based on" the evidence and the prior beliefs, as reflected by the fact that different beliefs would have been acquired if these factors had been different. More importantly, we can ask whether the bridge principles encoded in the prior probabilities are themselves reasonable. Is the agent justified to treat E_S as supporting H?

This question can be addressed from different angles. Obviously, we can ask whether the relevant prior probabilities were in turn supported by even earlier probabilities and intermediate sensory evidence. But this only pushes the question further back. Suppose

11 Arguably, various more specific psychological facts can also be explained from this perspective. For example, [Gershman et al. 2012] argue that if our visual system employs a form of Monte Carlo simulation to (approximately) compute conditional probabilities, then we should expect that when processing ambiguous figures like the duck rabbit, one interpretation of the figure should usually dominate, although the dominant interpretation should occasionally flip without any change in the perceptual input – which is of course precisely what we find (see also [Weiss et al. 2002], [Sanborn et al. 2010], [Gershman and Daw 2012]).

the agent's beliefs evolved from an *ultimate prior* probability function by successive conditionalisation on sensory evidence. Does this automatically make their belief state epistemically rational? Arguably not. There are substantive, non-formal constraints on beliefs and how they change through perceptual experience. Absent unusual background beliefs, a visual experience of rain should not make you confident that Tycho Brahe was poisoned by Johannes Kepler; observation of a million green emeralds should not make you certain that all unobserved emeralds are blue. In our Bayesian model, these norms map onto constraints on prior probabilities conditional on imaginary propositions.

Where do these norms come from? It is doubtful that they could be defended by non-circular a priori reasoning that would convince a skeptic who maintains that our experiences shed no light on what's going on in the external world or that any response to experience is as good as any other. Perhaps some such norms have to be taken as basic. Or perhaps they can be explained as (in some sense) constitutive of the relevant intentional states: on this view, what makes it true that a given state is a belief that it is raining is in part that it is normally caused by perceptions of rain.

We might also try to vindicate at least some constraints on ultimate priors by objective, external correlations. Suppose that sensory stimulus S is triggered mostly under external circumstances C, and robustly so. Then we might say that *S objectively supports C*. More generally, if circumstances C obtain in x percent of cases in which S is triggered, we might say that S objectively supports C to degree x. Accordingly, we could say that a subject is objectively justified in assigning conditional credence x to C given E_S if x matches the degree of objective support.

These issues and options are familiar from contemporary discussions in epistemology. The Bayesian model does little to resolve them. What it does is provide a credible background story. The present flavour of hardcore empiricism does not presuppose an outdated, 18th century view of perception and the mind. On the contrary, it goes very naturally with 21st century cognitive science and artificial intelligence.

It also provides a response to many objections that have been mounted against traditional flavours. For example, our Bayesian model is not committed to an ontology of sense data or irreducible phenomenal properties. A robot that follows the Bayesian model could live in a completely physical world. Relatedly, our model makes no claims about what we *see* (or hear or taste), or about what we are *directly aware of* in experience. Surely what we see are in general such things as trees and tables and tigers. The Bayesian model does not postulate that we also, or primarily, see non-physical ideas, impressions or sense-data. It also doesn't contradict the claim that perceptual experiences have a kind of content that rules out possibilities about the world, as reflected in our intuitions about veridicality.

The Bayesian model easily accommodates the holism of confirmation. Whether sensory information E_S supports a genuine hypothesis H about the world generally depends on the agent's background beliefs. Depending on the prior probabilities, the very same experience can rationally lead to very different beliefs. There is no once-and-for-all right or wrong interpretation of sensory signals. The only propositions that are directly and unrevisably supported by sensory experiences are imaginary propositions without real empirical content.

Many recent discussions of empiricism and foundationalism assume the demanding conception of 'belief' tied to linguistic assertions and corresponding acts of inner speech. This renders the foundationalist picture very unappealing. As Sellars [1956] pointed out, it is implausible that merely having a perceptual experience is sufficient to acquire beliefs in this sense, as one could plausibly have the experience without having the relevant linguistic or conceptual competence. It is also patently implausible that in any perceptual experience, we first become certain of a special "observation sentence" from which we then deduce other, perhaps probabilistic, statements about the world. The Bayesian model is not committed to these implausible claims. Conditionalisation is not an inference, with premises and conclusion, nor is it supposed to be a conscious, deliberate activity. Crucially, you don't need to have words that capture the imaginary content of all your perceptions.[12]

If 'belief' and its cognates are used in the demanding, intellectualist sense, it is hard to explain the epistemic impact of perceptual experiences. Since perceptions don't seem to have the required sentence-like, "conceptual" content, it looks like they can't support or justify beliefs with that kind of content; the link between perception and belief would then be merely causal and thus outside the domain of epistemology. Once we broaden our conception of intentionality in the manner envisaged by Ramsey, this problem (a version of "Sellars's dilemma") disappears. One can model perceptions as content-bearing states without implying that they somehow involve words or concepts.

In the previous section, I suggested that one might nevertheless want to restrict terms like 'belief' to attitudes whose contents are real propositions. From this perspective, the Bayesian model isn't a flavour of traditional empiricist foundationalism. It rather vindicates Jeffrey's own "softcore empiricism". It offers a systematic account of how perceptions affect credences without rendering anything certain, and without assuming any fixed probabilistic connection between experiences and propositions about the world.

As an epistemology, Jeffrey's radical probabilism is as unsatisfactory as coherence theories of justification in traditional epistemology. An adequate epistemology should have something to say on how perceptions affect rational beliefs. The Bayesian model offers an answer, whether or not we count attitudes towards imaginary propositions as

12 We often convey information about relevant features of an experience by reference to conditions that normally cause experiences with those features, as when we say that a stick looks broken, or feels cold. But of course these judgments are not infallible, nor do they capture the proposition we actually conditionalise on.

'belief'.

5 Puzzles of consciousness

Is there any introspectible trace of our alleged certainty in imaginary propositions? The model does not require it, but arguably there is. Consider your present perceptual experience. Are there any possibilities you can intuitively rule out in virtue of having this experience? Don't think of this as an attitude towards a sentence. Rather, imagine different ways things could be and ask yourself whether any of them can be ruled out given your experience. For example, consider a scenario in which you are skiing – a normal skiing scenario, without pervasive hallucinations, rewired brains, evil demons or the like. It could be a real situation in the past, if you ever went skiing. Your experiences in this situation are completely unlike your actual present experiences. (I trust you are not reading this paper while skiing.) In the skiing scenario, you see the snow-covered slopes ahead of you, feel the icy wind in your face, the ground passing under your skis, and so on. What is your credence that this situation is actual right now? Arguably zero. In general, when we have a given experience we can rule out any situations in which we have a suitably different experience. Or so it certainly seems. That is why skeptical scenarios almost always hold fixed our experiences and only vary the rest of the world.

These intuitions put pressure on physicalist accounts of experience. If experiences are brain states, and we can always rule out situations in which we have different experiences, it would seem to follow that merely in virtue of being in a given brain state we can rule out situations where we are in different brain states. That seems wrong. As Lewis [1995: 329] says: "Making discoveries in neurophysiology is not so easy!" Lewis concludes that as a physicalist, he has to reject the folk psychological *Identification Thesis*, that when we have an experience of a certain type, we can rule out possibilities in which we have experiences of a different type.[13]

Our Bayesian model shows that physicalism is in fact compatible with (a version of) the Identification Thesis. Suppose experiences of reading and experiences of skiing are associated with different imaginary propositions on which you conditionalise when you have the relevant experience. Then there is a sense in which you can indeed rule out the skiing scenario merely in virtue of having the reading experience. In your extended doxastic space, the skiing scenario isn't just characterised by its physical or functional properties. You cannot conclusively rule out any physical hypothesis about your experience. The scenarios you can rule out in your extended doxastic space are ruled out not by their physical features, but by their imaginary features.

13 According to Lewis, the Identification Thesis as just stated is equivalent to the claim that we "know the essence" of the relevant experience type; but this formulation is at best misleading (see [Stoljar 2009]).

These considerations suggest that the Bayesian model might also explain other facts about our epistemic situation that are widely thought to pose a threat to physicalism. To an agent with the kind of architecture I have outlined, it will seem as if there are facts that go beyond ordinary, physical facts about the world, even if she lives in a completely physical world. Since there is no logical entailment from real propositions to imaginary propositions, physical information about the world will always leave questions open in the agent's doxastic space.

Consider Mary (from [Jackson 1982]), who has learned all physical facts about colours and colour vision without having seen colours. If Mary's visual system implements the Bayesian model, then her doxastic space will contain an extra dimension, and her physical knowledge won't help her to rule out various possibilities along this dimension. In particular, if she is finally handed some coloured chips (without being told their colours), she can entertain different possibilities about what will happen when she will look at green grass, for she still won't know which type of imaginary proposition will then be true. But this will be so even if Mary lives in a completely physical world. Hence her epistemic situation gives us no reason to believe that physics is incomplete. Similarly, if Q is an imaginary proposition associated with red experiences, and P is the totality of all physical truths, then both $P \,\&\, Q$ and $P \,\&\, \neg Q$ will be a priori conceivable for Bayesian agents (see [Chalmers 2009]), but again this doesn't warrant any conclusions about the completeness or incompleteness of physics.

Dualists about the mind are under pressure to explain why specific phenomenal properties invariably occur under certain physical conditions, perhaps by postulating a large number of fundamental psycho-physical laws. The Bayesian model immediately explains this correlation, for it entails that the imaginary proposition a Bayesian agent learns on a given occasion is determined by her biological and physical features at that time – in the simplest case, by her sensory stimuli. So the model explains not only the appeal of dualism, but also that of "a posteriori physicalism": the idea that phenomenal states really are identical to physical states, although the identity is not discoverable by armchair reflection. (And obviously, the model has no problem accommodating the causal closure of physics.)

A subtle question is whether the intuitive judgments that appear to support dualism are mistaken on our model. It depends on whether we evaluate these judgments relative to the agent's extended doxastic space or relative to the space of real propositions. Assuming physicalism, there are no *real* questions whose answer is not entailed by broadly physical truths. Mary can't be ignorant of any real facts once she knows all the physical facts. If we only consider real propositions, Mary doesn't acquire any new knowledge when she first sees colours. But nor does she merely acquire new abilities or new concepts. What happens is that her extended subjective belief function gets conditionalized on an imaginary proposition – which is why it appears to Mary (and to us, when we imagine her situation) that she acquires genuine knowledge.

On the other hand, relative to her own doxastic space, Mary does learn something new when she first sees colours. She can rule out previously open possibilities, and it makes sense to classify this as an acquisition of knowledge.

Here the analogy to self-locating beliefs may help. If you already knew every truth about the universe from a God's eye perspective and then find out who you are and what time is now, do you acquire new knowledge? In a sense, no: there is no objective proposition about the world you previously failed to know. On the other hand, in your own doxastic space, finding out what time it is excludes genuine possibilities, albeit not objective possibilities about the world. What you learn is a *centred proposition*. Centred propositions are true or false only relative to further parameters beyond an ordinary possible world. Since these parameters are provided by the context in which an agent believes or disbelieves the proposition, her attitude can be evaluated as true or false (simpliciter, or relative to a world alone). For example, if at noon an agent believes the centred proposition that it is noon, then her belief is true, and may qualify as knowledge. Similarly for imaginary and complex propositions. These, too, are true or false only relative to a possible world (or centred world) plus something else. That *something else*, a sense datum, is effectively provided by an agent's present sensory signals, so again her beliefs in such propositions can reasonably be classified as true or false. And so it makes sense to say that Mary's new beliefs about colours are true, and amount to knowledge.

What is important in both cases is that the objective picture of the world predicts the extra possibilities that open up from the subjective point of view. If our cognitive architecture approximates the Bayesian model, then this will be entailed by objective physical facts. The physical facts will reveal that our doxastic space has an imaginary dimension, just as it has a self-locating dimension. This provides an answer to the "Master Argument" in [Chalmers 2007]. The argument, in a nutshell, is that any physical information about us equally applies to zombies and thus can't explain our special epistemic situation with respect to phenomenal consciousness. On the present proposal, broadly physical information about our cognitive architecture implies that our epistemic space has an imaginary dimension, and that our beliefs evolve by conditionalising on propositions in this dimension. Perhaps it is a priori conceivable (in our extended epistemic space) that we are zombies for whom all this happens "in the dark". But these zombies will share a lot of our epistemic situation. In particular, the physical facts ensure that zombie Mary gains cognitively significant information when she first sees colours, just like Mary.[14]

14 Chalmers's argument is directed against the "phenomenal concepts strategy" which tries to explain the epistemic gap between physical and phenomenal truths as an artifact arising from the nature of our phenomenal concepts: these are supposed to pick out physical states in a special way that renders the connection to ordinary physical information epistemically opaque. The main challenge for this

To be sure, many aspects of our conscious experience are not explained by the Bayesian model from section 2. I never even mentioned consciousness when I presented the model. Presumably the events we call 'conscious' play a special role in our cognitive architecture, related to the fast-track processing about which I speculated in section 3. I have nothing substantial to contribute to this question. In the terminology of [Chalmers 1995], the present model only solves the "hard problem" of consciousness, not the "easy problems".

I want to close with a few remarks on another puzzle about consciousness that has not received much attention in the literature ([Latham 2000] is an exception). The puzzle is the apparent *fit* between the phenomenal character of mental states and their functional role. To see what I mean, compare again the skiing scenario on the one hand with your actual present situation on the other. Both situations come with a distinctive phenomenal character. For the skiing scenario, this involves the phenomenology of feeling the wind, seeing the slopes, moving your legs, and so on. My claim is that this phenomenal character goes well with the external circumstances that caused it and the behaviour it causes. Imagine a world where the phenomenal characters are swapped, where ordinary skiing events are associated with the actual phenomenology of sitting at a desk and reading a paper, and vice versa. This would be a world where phenomenal character doesn't fit functional role.

Are "inverted qualia" worlds like this conceivable? Not if the phenomenal truths are a priori entailed by broadly physical truths. But many philosophers – physicalists and dualists alike – deny the thesis of a priori entailment. They hold that there is an epistemic gap between the physical and the phenomenal. This suggests that worlds with thoroughly inverted qualia should be epistemically possible.[15] Epistemically speaking, it is then just a coincidence that in our world phenomenology nicely fits functional role. For all we know a priori, it could have been that skiing experiences are associated with the phenomenology of reading philosophy papers. Or it could have been that everyone's phenomenology is running two hours late so that, when people eat breakfast and read the newspaper in the morning, they have the experience of still sleeping; when they have started working, they have the experience of eating breakfast and reading the newspaper, and so on. How convenient that we don't live in a world like that! (If indeed we don't.

strategy is to explain how this opaque reference to physical states is supposed to work (see [Chalmers 2007], [Levine 2007]). The present proposal does not need any claim about reference.

15 Strictly speaking, one could deny that the phenomenal is entailed by the physical but also deny the coherence of the described scenario. The idea would be that there is *partial* entailment from the physical to the phenomenal: given a state's physical and functional properties, one can a priori rule out many candidate phenomenal properties; the entailment is partial because more than one candidate is left standing. However, most philosophers who believe in an explanatory gap believe that the gap is fairly wide, so that physical information entails very little about phenomenal character. As long as the gap is sufficiently wide, we can construct strange inversion scenarios, even if not the exact scenario from above.

For can we really be sure?)

This is the puzzle. Here is the solution. In our extended doxastic space, imaginary propositions are compatible with many, perhaps all, real propositions. There are points in our doxastic space where the imaginary proposition actually associated with skiing stimulations is associated with reading stimulations. On the other hand, in order for perception to provide us with information about the world, there must be strong a priori constraints on the interpretation of sensory signals, i.e. on the probabilities of real propositions conditional on imaginary propositions. Absent unusual background information, certain signals must be regarded as evidence for the presence of light, of such-and-such shapes, and so on. These connections can change through experience, but the functioning of our perceptual system demands that we give low a priori probability to possibilities where a given type of experience is associated with unusual causes and effects. The inverted qualia worlds just described are extreme scenarios of this type. Our Bayesian model predicts that they must have negligible probability. They are *almost* a priori ruled out.

References

David Chalmers [1995]: "Facing Up to the Problem of Consciousness". *Journal of Consciousness Studies*, 2(3): 200–219

— [2007]: "Phenomenal Concepts and the Explanatory Gap". In T. Alter and S. Walter (Eds.) *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism,* Oxford: Oxford University Press

— [2009]: "The Two-Dimensional Argument Against Materialism". In Brian McLaughlin (Ed.) *Oxford Handbook to the Philosophy of Mind,* Oxford University Press

David Christensen [1992]: "Confirmational Holism and Bayesian Epistemology". *Philosophy of Science*, 59(4): 540–557

Persi Diaconis and Sandy L. Zabell [1982]: "Updating Subjective Probability". *Journal of the American Statistical Association*, 77: 822–830

Jacob Feldman [2013]: "Tuning Your Priors to the World". *Topics in Cognitive Science*, 5: 13–34

Hartry Field [1978]: "A Note on Jeffrey Conditionalization". *Philosophy of Science*, 45(3): 361–367

Daniel Garber [1980]: "Field and Jeffrey Conditionalization". *Philosophy of Science*, 47(1): 142–145

Samuel Gershman and Nathaniel D. Daw [2012]: “Perception, action and utility: the tangled skein”. In M. Rabinowich, K. Friston and P. Varona (Eds.) *Principles of Brain Dynamics: Global State Interactions,* Cambridge (MA): MIT Press, 293–312

Samuel J. Gershman, Edward Vul and Joshua B. Tenenbaum [2012]: “Multistability and Perceptual Inference”. *Neural Computation*, 24: 1–24

James Hawthorne [2004]: “Three Models of Sequential Belief Updating on Uncertain Evidence”. *Journal of Philosophical Logic*, 33(1): 89–123

Yanping Huang and Rajesh PN Rao [2011]: “Predictive coding”. *Wiley Interdisciplinary Reviews: Cognitive Science*, 2(5): 580–593

Frank Jackson [1982]: “Epiphenomenal Qualia”. *Philosophical Quarterly*, 32: 127–136

Richard Jeffrey [1965]: *The Logic of Decision*. New York: McGraw-Hill

— [1975]: “Carnap’s Empiricism”. In G. Maxwell and R.M. Anderson (Eds.) *Induction, Probability, and Confirmation,* vol 6. Minneapolis: University of Minnesota Press, 37–49

— [1984]: “Bayesianism with a Human Face”. In J. Earman (Ed.) *Testing Scientific Theories,* Minneapolis: University of Minnesota Press, 133–156

— [1988]: “Conditioning, kinematics, and exchangeability”. In B. Skyrms and W.L. Harper (Eds.) *Causation, chance and credence,* Dordrecht: Kluwer, 221–255

— [1992]: *Probability and the Art of Judgment*. Cambridge: Cambridge University Press

Daniel Kersten, Pascal Mamassian and Alan Yuille [2004]: “Object perception as Bayesian inference”. *Annual Review of Psychology*, 55: 271–304

David C. Knill and Alexandre Pouget [2004]: “The Bayesian brain: the role of uncertainty in neural coding and computation”. *Trends in Neurosciences*, 27(12): 712–719

Noa Latham [2000]: “Chalmers on the addition of consciousness to the physical world”. *Philosophical Studies*, 98: 71–97

Joseph Levine [2007]: “Phenomenal Concepts and the Materialist Constraint”. In T. Alter and S. Walter (Eds.) *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism,* Oxford: Oxford University Press

David Lewis [1995]: “Should a Materialist Believe in Qualia?” *Australasian Journal of Philosophy*, 73: 140–144

Frank Ramsey [1931]: "Truth and Probability (1926)". In R.B. Braithwaite (Ed.) *Foundations of Mathematics and other Essays,* London: Routledge & P. Kegan, 156–198

Stuart J. Russell and Peter Norvig [2004]: *Artificial Intelligence: A Modern Approach.* Cambridge (MA): MIT Press, 2nd edition

A. Sanborn, T. Griffiths, D. Navarro, S. To and A. Sanborn [2010]: "Rational approximations to rational models: Alternative algorithms for category learning". *Psychological Review*, 117: 1144–1167

Wolfgang Schwarz [2012]: "Changing Minds in a Changing World". *Philosophical Studies*, 159: 219–239

Wilfrid Sellars [1956]: "Empiricism and the Philosophy of Mind". In H. Feigl and M. Scriven (Eds.) *Minnesota Studies in the Philosophy of Science,* vol 1. Minneapolis: University of Minnesota Press, 253–329

Brian Skyrms [1980]: "Higher Order Degrees of Belief". In D.H. Mellor (Ed.) *Prospects for Pragmatism,* Cambridge: Cambridge University Press

— [1984]: *Pragmatics and Empiricism.* Yale: Yale University Press

— [1990]: *The Dynamics of Rational Deliberation.* Cambridge (Mass.): Harvard University Press

Robert Stalnaker [1984]: *Inquiry.* Cambridge (Mass.): MIT Press

Daniel Stoljar [2009]: "The Argument From Revelation". In D. Braddon-Mitchell and R. Nola (Eds.) *Conceptual Analysis and Philosophical Naturalism,* Cambridge (Mass.): MIT Press, 113–138

Carl G. Wagner [2002]: "Probability Kinematics and Commutativity". *Philosophy of Science*, 69(2): 266–278

— [2013]: "Is Conditioning Really Incompatible with Holism?" *Journal of Philosophical Logic*, 42: 409–414

Jonathan Weisberg [2009]: "Commutativity or holism? A dilemma for conditionalizers". *The British Journal for the Philosophy of Science*, 60(4): 793–812

Y. Weiss, E. P. Simonvelli and E. H. Adelson [2002]: "Motion illusions as optimal percepts". *Nature Neuroscience*, 5: 598–604