Imaginary Foundations: Sense data in Bayesian epistemology*

Wolfgang Schwarz
Draft, 6 May 2018

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 may 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 proposal might also explain 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

* Ancestors of this paper were presented in Bielefeld (2006), Canberra (2006), Cologne (2010), Melbourne (2013), Leeds (2014), and Bochum (2017). I thank the audiences at these events for helpful feedback. Special thanks to David Chalmers, Colin Klein, Christian Nimtz, Mark Sprevak, and Daniel Stoljar for detailed comments on earlier versions.

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 problems. Aren't we often ignorant or 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?

In response, one might be tempted by alternative ways of defining the information provided by an experience. For example, if my experience is in fact caused by rain, and experiences of the same type are caused by rain across a variety of nearby worlds, then there is a good (causal) sense in which the experience "carries the information" that it is raining. But it is not clear how this sense of information bears on what I should belief in light of the experience. After all, I should *not* become absolutely certain that it is raining. My miniscule credence in the vodka hypothesis should not decrease. If I had reason to believe that the rain is staged, I should even remain confident that it is not raining.

So perhaps we should drop the assumption that perceptual experiences put us in a position to conclusively exclude possibilities. My experience, perhaps together with my background beliefs, merely allows me to conclude tentatively and defeasibly that it is raining. In general, on this view, 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 idea in the framework of Bayesian epistemology. Here we assume that beliefs come in degrees that satisfy the mathematical conditions on a probability measure. How should these probabilities change under the impact of perceptual experience? Classical Bayesianism suggests the following answer. For each type of perceptual experience there is a proposition E such that, whenever a rational agent has the experience, her new probability equals her previous probability conditional on E; that is, for all A,

$$P_{new}(A) = P_{old}(A/E) = P_{old}(A \wedge E)/P_{old}(E), \text{ provided } P_{old}(E) > 0.$$

Here, P_{new} is said to come from P_{old} by *conditionalizing on* E.[1] Since $P_{old}(E/E) = 1$, the new probability of E is 1. So E can hardly be an ordinary proposition about the world. Looking through the window does not make me certain that it is raining, nor should it. Again, we seem forced to postulate a mysterious realm of sense-datum propositions.

To avoid commitment to such propositions, Richard Jeffrey developed what he called *radical probabilism* as an alternative to the classical Bayesian picture (see [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

1 The conditional probabilities $P_{old}(A/E)$ are often computed via Bayes' Theorem, which is why conditionalization is also known as *Bayes' Rule*.

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 has become 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 probabilities 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. It 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 conditionalization* or *probability kinematics* or *generalised conditionalization* – but the last name is a little misleading, as we will see in a moment.[2]

At first glance, Jeffrey's model seems to deliver just what we were looking for. Instead of assuming that each type of perceptual experience is associated with a sense-datum proposition E rendered certain by the experience, we only need to assume that there is some assignment of probabilities $x_1, \ldots, x_n$ to the elements of 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 conditionalization.

More concretely, we might assume that for every perceptual experience there is a proposition E that captures how the experience intuitively represents the world as being. We do not require agents to become absolutely certain of E when they have the experience. Instead, we might say that they should assign some intermediate credence x (maybe 0.95) to E, and consequently $1-x$ to $\neg E$. The complete update is then determined by the following special case of Jeffrey's rule:

$$P_{new}(A) = P_{old}(A/E) \cdot x + P_{old}(A/\neg E) \cdot (1 - x).$$

We would still need 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 – not if experiences are individuated by their physiology or phenomenology.[3] For then the rational response to a given experience should depend on the agent's background information. Your new beliefs about

2 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].

3 Experiences can of course be typed in other ways. For example, we might say that two experiences are of the same type iff they lead to the same rational posterior beliefs. The difficulties I am going to discuss then resurface as the problem of determining when two experiences are of the same type.

the colour of the tablecloth, for instance, should be sensitive to background beliefs about the colour of other tablecloths in the house. My belief about the weather should be sensitive to background beliefs about whether or not people are filming a rain scene outside my window.

So we cannot associate experience types with fixed posterior probabilities $x_1, \ldots, x_n$ over fixed propositions $E_1, \ldots, E_n$. We must also take into account the agent's previous probabilities P_{old}. But how does a given type of experience, together with an agent's previous probabilities P_{old}, 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 question is sometimes called the *input problem* for Jeffrey conditionalization. It was first raised by Carnap in his 1957 correspondence with Jeffrey (published in [Jeffrey 1975]). Here Carnap reports that he had himself toyed with the idea of relaxing the classical 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.[4]

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'd had two equally inconclusive but very different experiences of the cloth, from different angles perhaps, 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 each experience together with a specification of the 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.[5]

4 The input problem also arises for standard ("strict") conditionalization if we don't assume that the proposition on which agents conditionalize is determined by the relevant experience. For example, [Skyrms 1980] points out that the effect of Jeffrey conditionalization can be mimicked by strict conditionalization on certain propositions about (posterior) degrees of belief, so that what you learn for certain in the tablecloth scenario is (say) that you have become 60 percent confident that the tablecloth is green. The input problem now turns into the question how your experience together with your prior beliefs determines that your new credence in the tablecloth being green should be 0.6.

5 The present point is inspired by [Garber 1980], where it is used to argue against a particular answer to the input problem suggested in [Field 1978]. [Hawthorne 2004] presents a model that gets around the problem by making the input parameters to Jeffrey conditionalization depend not only on present experience and old probabilities, but also on earlier experiences; the dynamics of rational credence

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 $x_1, \ldots, x_n$ over a partition $E_1, \ldots, E_n$ such that probabilities conditional on the partition cells are preserved. All Jeffrey 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 conditionalization as the second norm of Bayesian epistemology. His alternative threatens to collapse 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. There are substantive norms on how one's beliefs may change through perception. For example, when a chemist uses a litmus strip 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. Likewise, 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 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 the rational impact of a perceptual experience on an agent's beliefs is not determined by the nature of the experience, her previous beliefs, and the environment.* Something else plays a role. We need to know what it is and how it works.

---

is thereby rendered unattractively non-Markovian. Further (though related) 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).

2 Armchair robotics

To make progress on the problem 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 surroundings – 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 activities in the sense organs?

A sense organ is a physical device whose internal state systematically and reliably varies with 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 light 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 signals produced by the sense organ – i.e., to different patterns of electrochemical activity in the "output" wires of the device.

It would be convenient if one could read off the exact colours, shapes, and spatial arrangement of objects in the environment from the signal produced by the robot's sense organ. In practice, this is not possible, because different configurations of the environment lead to the very same activation of photoreceptors and thus to the very same sensory signal: from certain angles, a three-dimensional cube can cause the very same signal as a two-dimensional picture of a cube; a small cube at close distance can cause the very same signal as a larger cube further away; a convex shape with light from above can cause the same signal as a concave shape with light from below; a red cube under white light can cause the same signal as a white cube under red light; and so on.

So the functional architecture of a sense organ only determines, for each sensory 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 conditionalization, where the new probabilities $x_1, \ldots, x_n$ over $E_1, \ldots, E_n$ might reflect something like the ecological relative frequency or objective chance with which the conditions obtain when the signal is produced.[6]

6 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

But that is still not an optimal solution. The new probability assigned to the E_i's should be sensitive not only to the sensory signal (equivalently, to the upstream activation of photoreceptors) 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.

A better idea, which gets us closer to actual approaches in Artificial Intelligence (see e.g. chs. 15 and 17 in [Russell and Norvig 2010]), is to fix not the probability of E_i given S, but the inverse probabilities of S given E_i. That is, let's endow our robot with a "sensor model" that defines a probability measure π over possible signals S conditional on possible world states E_i. The new probabilities over the world states can then be computed by a variant of Bayes' Theorem:

$$P_{new}(E_i) = \frac{\pi(S/E_i)P_{old}(E_i)}{\sum_j \pi(S/E_j)P_{old}(E_j)}.$$

Note that $P_{new}(E_i)$ is sensitive to $P_{old}(E_i)$, as desired.

A minor problem with this approach is that the sensor model should arguably not be fixed once and for all either. To fix this, let's merge the "sensor model" π into the robot's main probability function P. That is, we extend the domain of P by the set of possible sensory signals, and replace $\pi(S/E_i)$ by $P_{old}(S/E_i)$. Intuitively, we assume that our robot has opinions about what kinds of signals it is likely to receive in what kinds of environments. These opinions can themselves be sensitive to sensory input.

If we replace $\pi(S/E_i)$ in the above variant of Bayes' Theorem by $P_{old}(S/E_i)$, then $P_{new}(E_i)$ is simply $P_{old}(E_i/S)$. Moreover, Jeffrey conditionalizing on a partition $E_1, \ldots, E_n$ whose new probabilities $x_1, \ldots, x_n$ are given by $P_{old}(E_1/S), \ldots, P_{old}(E_n/S)$, respectively, is equivalent to strict conditionalization on S. So we might as well bypass the evidence partition $E_1, \ldots, E_n$ and simply say that for all A,

$$P_{new}(A) = P_{old}(A/S).$$

That is, the new probability of any proposition A is the old probability of A conditional on the current sensory signal. Our robot has become a classical ("strict") conditionalizer. What it conditionalizes on when it receives a sensory signal is not any of the associated propositions $E_1, \ldots, E_n$ about the environment, but rather the signal itself.

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 baseball bat. So we might set $x_{n+1} > 0$ and adjust the other probabilities accordingly.

There's something odd about this approach. It looks like our robot must now have well-defined subjective probabilities over sensory signals, i.e. over the occurrence of complicated electrochemical events at the interface of its sense organs. One might have thought that a robot in search of mushrooms wouldn't need to be trained in electrochemistry, and that it wouldn't need to have perfect knowledge about the internal workings of its perceptual system. Indeed, a little reflection makes clear that this is not required.

Suppose the robot's probabilities are originally defined only for certain propositions $\mathbf{R}$ about the macroscopic environment. The above considerations suggest that we need to extend the domain of the probability function by further elements $\mathbf{I}$ such that when a signal S arrives, the robot conditionalizes on a corresponding proposition $\rho(S) \in \mathbf{I}$, where ρ is some function mapping distinct signals to distinct elements of $\mathbf{I}$. The elements of $\mathbf{I}$ may therefore "represent" or "denote" electrochemical events in some causal sense, but this need not in any way be transparent to the robot or reflected in its probability space. For example, if at some point we wanted to replace the robot's photoreceptors with ones that produce different electrochemical outputs, we would need to adjust the mapping ρ, but we would not have to to change the content of the robot's database. Similarly, if we eventually wanted to train our robot in electrochemistry, the elements of $\mathbf{I}$ would still not need to stand in interesting set-theoretic relationships to the electrochemical propositions in the robot's belief space.

To understand what I'm trying to suggest, it may help to imagine that our robot stores information in the form of English sentences, so that its database pairs sentences like 'there is a mushroom to the left' with numbers between 0 and 1. If we restrict the database language to ordinary sentences about the robot's macroscopic environment, we will run into problems when we want to specify how the database should be updated in response to activity in the robot's sense organs. To optimally deal with sensory input, I suggest, we need to extend the robot's probability space by new sentences such that whenever a sensory signal arrives, the robot becomes certain of one of these sentences. But there is no good reason why these sentences must be correct and detailed descriptions of the relevant electrochemical signal. In principle, the update works just as well if the new sentences are bare tags, 'A', 'B', 'C', etc.

I will have more to say on what, if anything, the elements of $\mathbf{I}$ represent, and whether the robot's probabilities over $\mathbf{I}$ should be understood as degrees of belief. Formally, the required extension of a probability measure is a straightforward product construction. Take the simplest case where everything is finite. Let $\mathbf{R}$ be the set of propositions about the environment on which we want the robot to have an opinion. Probability theory requires that $\mathbf{R}$ is a Boolean algebra; so we can identify each proposition in $\mathbf{R}$ with a set of "possible worlds": the atoms of the algebra. Now let $\mathbf{I}$ be an arbitrary set disjoint from $\mathbf{R}$ such that there is a one-one correspondence between $\mathbf{I}$ and the signals the robot

can receive.[7] Each pair $\langle w, e\rangle$ of a possible world w and an element e of $\mathbf{I}$ 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 the robot's extended probability measure assigns a value a *(complex) proposition*. The members of $\mathbf{R}$ are *real propositions*; subsets of $\mathbf{I}$ are *imaginary propositions* – in analogy to complex, real, and imaginary numbers, and to highlight the fact that imaginary propositions do not have to be understood as genuine propositions about a special subject matter. Individual members of $\mathbf{I}$ I will call *sense data*, since their role in the present model in some ways resembles the role of sense data in the classical empiricist model of perception (see section 4 below). In the robot'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.

In some respects, the construction of complex propositions is analogous to a popular construction of *centred propositions* in the modelling of self-locating beliefs. Arguably, our doxastic space contains not only propositions about the universe as a whole, but also propositions about our own current place in the world: an agent might know every truth about the world from a God's eye perspective and still be ignorant about who they are or what time is now (see e.g. [Lewis 1979]). 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 that fixes an individual and a time in w. In the resulting doxastic space, an objective proposition A about the world is then re-identified with the set of "centred worlds" $\langle w, c\rangle$ whose possible-world coordinate w lies in A.

In fact, there are good reasons to make the propositions in our robot's doxastic space centred as well. Imagine the robot is moving towards a wall. At time t_1 it receives a signal S which (by the robot's lights) indicates that the wall is about 5 metres away. A little later, at t_2, the robot receives another signal S' indicating that the wall is about 4 metres away. At this point, we don't want the robot to conclude that the wall 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 wall was 5 metres away *at the time*, which is perfectly compatible with the distance now being 4 metres.

Here is how the model I have outlined could be adjusted to accommodate the passage of time. (These details will not be important for what follows.) First, we make the objects of probabilities centred. So probabilities are now defined over a three-fold product $\mathbf{R} \times \mathbf{I} \times \mathbf{T}$ (or a sub-algebra of that product), where $\mathbf{T}$ is a suitable set of relative time indices. Assuming for simplicity that time is linear and discrete, we might identify $\mathbf{T}$ with the set of integers, interpreting 0 as *now*, 1 as *the next point in the future*, and so on.

7 $\mathbf{I}$ is an "arbitrary" set because the identity of its members is irrelevant to the functional specification of our robot. In this respect, the framework of probability theory is a little artificial, since it forces us to make a choice.

When signal S arrives, the robot conditionalizes not on $\mathbf{R} \times \{\rho(S)\}$, as above, but on $\mathbf{R} \times \{\rho(S)\} \times \{0\}$ – intuitively, on the indexical proposition that $\rho(S)$ is true *now*. Such indexical beliefs must be updated constantly to keep track of the passing time. Thus at the next point in time, the robot's certainty of $\mathbf{R} \times \{\rho(S)\} \times \{0\}$ should have evolved into certainty of $\mathbf{R} \times \{\rho(S)\} \times \{-1\}$ – intuitively: $\rho(S)$ was true *one moment ago*. These updated probabilities are then conditionalized on the new evidence $\mathbf{R} \times \{\rho(S')\} \times \{0\}$. See [Schwarz 2017] for further details and motivation.

3 From the armchair to cognitive science

I have described a model of how subjective probabilities should change under the impact of sensory stimulation. The model requires an agent's doxastic space to be extended by an "imaginary" dimension whose points are associated with sensory signals in such a way that when a given signal arrives, the agent assigns probability 1 to the corresponding imaginary proposition; the probability of real propositions is then adjusted in accordance with their prior probability conditional on that imaginary proposition.

I already mentioned that the model is not really new. It closely resembles standard treatments in artificial intelligence. It might also be familiar from the neuroscience of perception, where similar models have proved a useful paradigm for studying our perceptual system (see [Yuille and Kersten 2006] for a survey). In these areas, the propositions on which an agent (or her perceptual system) is assumed to conditionalise are called 'sense data', 'raw data', or 'input strings', and people rarely pause to reflect on their representational features on what the postulated models imply for the epistemology of perception.

That is what I want to do. Suppose, in line with evidence from cognitive science, that our own cognitive system approximates something like the model I have described. What would that mean? Would it vindicate classical empiricist foundationalism? Would it provide an answer to the problems from section 1?

To begin, we need to clarify how the extended probability function that figures in the model should be interpreted. Does it represent the agent's degrees of belief? That depends on what we mean by 'belief'.

Philosophers often use terms like 'belief' or 'credence' in a demanding *intellectualist* sense tied to conceptual structure, conscious thought, and linguistic assertion. In that sense, it is doubtful whether cats, dogs, or robots have beliefs. The model I have described does not require that the relevant agents have a language, or that they store information in the form of "conceptually structured propositions" (whatever that might mean). So the extended probability function in the model arguably doesn't fit the job description for intellectualist degrees of belief.

The main alternative to the intellectualist conception of belief is a family of *functionalist*

conceptions on which belief and other intentional states are defined by their functional role. According to a crude version of this approach (still popular in some parts of economics), beliefs and desires are defined by an agent's behavioural dispositions: to have such-and-such beliefs and desire just means to be disposed to make such-and-such choices. Less crude versions of functionalism identify intentional states with causally efficacious internal states whose defining functional role links them not just to behavioural output but also to one another and to sensory input. Here, an agent has such-and-such beliefs just in case she is in some state or other that bears the right connection to sensory input, to behavioural output, and to other internal states similarly individuated by their functional role.[8]

On the functionalist approach, 'belief' is implicitly defined by a certain theory, or model. Different models may therefore define different notions of belief. Consider the classical Bayesian model of rational agents. The model assumes that agents have a credence and a utility function. What does it take for a lump of flesh and blood or silicon to have a credence and a utility function? On the functionalist conception, an agent has a certain credence and utility function just in case she is in some state that plays the role the model attributes to these functions. One aspect of that role links the state to the agent's actual and counterfactual choices: when facing a decision, standard Bayesianism assumes that an agent chooses an option that maximizes expected utility in light of her credence and utility function. Another aspect of the role describes how an agent's probabilities and utilities change over time. This is where the "input problem" arises. As we saw, it is hard to specify how an agent's probabilities should change in response to sensory stimulations (or perceptual experiences, if you want) if don't assume that these provide infallible access to relevant facts about the world. To get around this problem, I have suggested that we extend the domain of the probability function in the model by extra, "imaginary" elements associated with sensory signals in such a way that any given sensory stimulation leads to conditionalization on the associated imaginary element. What does it take for a concrete agent to have such an extended probability function? As before, the agent must be in some state or other that plays the functional role assigned to this function (to a sufficient degree of approximation).[9]

Even on the functionalist conception, however, there are limits to what one can sensibly

8 See e.g. [Lewis 1974], [Stalnaker 1984], or [Braddon-Mitchell and Jackson 1996] for classical expositions of this kind of functionalism in the philosophy of mind.

9 A full specification of the relevant functional role would need spell out other controversial details in the Bayesian model – for example, whether choices should maximise causal or evidential expected utility. The complete model should also include non-formal constraints on utilities and probabilities, for the reasons discussed in [Lewis 1974] and [Lewis 1983]: without such constraints, the model will plausibly allow for too many assignments of probabilities and utilities, many of which are far removed from what we would intuitively take to be the agent's beliefs and desires. I will turn to non-formal constraints on extended probabilities in a moment.

call 'belief'. If a certain role deviates too far from the role we ordinarily associate with the word 'belief', it would be better to use a different label. On these grounds, one might argue that an agent's probabilities over non-real propositions should not be called 'credences' or 'degrees of belief'. Intuitively, to have a belief is to represent the world as being a certain way. Beliefs can be true or false; partial beliefs can be accurate or inaccurate to various degrees. By contrast, in the model I have described, an agent's probabilities over imaginary propositions are not meant to represent information about the world; they can't assessed for accuracy. If anything, the non-real part of an agent's probability function seems to represent the agent's dispositions to change her (genuine) beliefs about the world in response to sensory input. For example, what makes it the case that our robot assigns greater probability to $A \,\&\, I_S$ than to $\neg A \,\&\, I_S$, where I_S is the sense datum $\rho(S)$ associated with signal S, is to a large part that receiving S will make the robot assign greater probability to A than to $\neg A$.

Of course, nothing of substance turns on whether we apply the label 'degree of belief' to an agent's entire extended probability function or only to the "real" part of that function. For the time being, let me adopt the more restrictive usage.

An agent's degrees of belief, on the restricted usage, do not evolve by strict conditionalization, since the relevant sense data are not in the domain of the belief function. Degrees of belief instead evolve by Jeffrey conditionalization. This is because the redistribution of an agent's probabilities over real propositions brought about by conditionalising the extended probability function on sense data can arguably always be modelled as an instance of Jeffrey conditionalization. Concretely, assume that $\{E_1, \ldots, E_n\}$ is some partition of real propositions that "screens off" the sense datum I_S from any other real proposition – meaning that $P_{old}(A/E_i \wedge I_S) = P_{old}(A/E_i)$ for any real A and any $E_i \in \{E_1, \ldots, E_n\}$. The agent's degrees of belief then change by Jeffrey conditionalization on $\{E_1, \ldots, E_n\}$, with the new credences $x_1, \ldots, x_n$ set by $P_{old}(E_1/S_I), \ldots, P_{old}(E_n/S_I)$, respectively. If the agent's real probability space is finite, there will always be some such partition (in the worst case, the partition of individual worlds). In the infinite case, we may have to invoke some generalisation of Jeffrey's formula like the one mentioned in footnote 2.

The model I have outlined therefore provides an answer to the input problem, at least for the functionalist conception of belief: we can explain how a given experience together with an agent's prior state determines the input parameters to a Jeffrey update – provided that the "prior state" includes an extended probability function of the kind I described. If we only look at the agent's prior probabilities over real propositions, the problem can't be solved: there is no fixed way in which perceptual experiences should affect an agent's beliefs about the world.

To illustrate how the present model gets around the problems from section 1, imagine our robot finds itself in the repeated tablecloth scenario. Let S be the perceptual signal the robot receives both times when it looks at the cloth in the dimly lit room, and let I_S

be the sense datum associated with S. Let's say the robot initially assigned probability $1/3$ to each of the possibilities $\{Red, Green, Blue\}$. Moreover, let's assume it assigned greater probability to I_S conditional on *Green* than to I_S conditional on *Red* or *Blue*. Concretely, let's assume that $P_{old}(I_S/Green) = 1/4$ and $P_{old}(I_S/Red) = P_{old}(I_S/Red) = 1/8$. As you can check, the robot's new probabilities over $\{Red, Green, Blue\}$ will then be $1/4$, $1/2$, $1/4$, respectively. At the same time, the robot's probability for I_S increases to 1. When the robot takes a second look at the cloth, we can assume that its extended probability measure assigns high probability to the hypothesis that it is going to learn I_S again. As a consequence, the second look at the tablecloth will barely affect the robot's beliefs about the cloth's colour.

Here I have made various assumptions about the robot's prior probabilities. The formal model alone does not guarantee sensible results. For example, if the robot assigned high prior probability to I_S-*now* conditional on *Red* and to I_S-*in-a-moment* conditional on *Blue*, the first look at the tablecloth might make it confident that the cloth is red and the second that the cloth is blue. In this respect, the present model is on a par with the classical model on which beliefs are assumed to evolve by strict conditionalization. That, too, leads to sensible posterior beliefs only if the agent starts out with sensible prior beliefs. I will return to the "problem of the priors" in the next section. First I want to discuss another important caveat.

Bayesian models of rationality are often highly idealised. The model I have described certainly is. Such idealised models can still be useful, and not just as normative ideals. Even if they don't fit all the phenomena, they can capture important patterns in the phenomena – central aspects of our psychology, ignoring friction and air resistance, as it were. On the other hand, it is also useful to study how reality deviates from the ideal. Here too one can often find interesting patterns. In fact, many peculiarities of our cognitive system can arguably be explained as consequences of the short-cuts evolution has taken to approximate the model I have described.

If we tried to actually build our robot, with its central database of probabilities updated by conditionalizing on sense data, we would quickly hit insurmountable problems. Conditionalizing a high-dimensional probability measure is a non-trivial, often intractable computational task. Researchers in cognitive science and theoretical computer science have suggested several tricks to make it tractable. For example, instead of computing exact conditional probabilities we could employ Monte Carlo sampling ([Griffiths et al. 2008]) or variational approximations ([Seeger and Wipf 2010]). Restricting the mathematical form of prior probabilities also proves useful in this context. Various ideas from predictive coding could be used to exploit regularities in sensory signals ([Clark 2013]). Decomposing an agent's ("joint") probability measure into probabilistically independent components, perhaps in the structure of a Bayes net, can also dramatically improve computational

tractability ([Pearl and Russell 2001]).[10] For evidence that our cognitive system employs these and other tricks to approximate our simple Bayesian ideal, see e.g. [Weiss et al. 2002], [Vul et al. 2009], [Sanborn et al. 2010], [Gershman and Daw 2012], [Gershman et al. 2012], [Howhy 2014], [Griffiths et al. 2015].

Another idea that could potentially help make our model tractable is to drop the assumption of precise probabilities. Instead of storing a precise probability function, it might be easier to merely store certain constraints on probabilities: that P is more probable than Q, that R is probabilistically independent from S, and so on. The agent's doxastic state would then be represented by a whole set of probability functions: all those that meet the constraints ([Jeffrey 1984]). On closer inspection, however, it is not clear whether this would improve computational tractability, since standard methods for updating a probability function (such as Monte Carlo sampling) do not straightforwardly generalise to sets of probability functions (see e.g. [Zhang et al. 2013]).[11] On the other hand, there are independent reasons to drop the assumption of precise probabilities. Several authors have argued that even ideal agents should not have precise credences (see e.g. [Joyce 2005], [Sturgeon 2008]). These arguments naturally carry over to an agent's extended probability function, suggesting that for many sense data S and real propositions A, rational agents should not assign a strict probability to S given A. Evaluating these arguments is beyond the scope of the present essay, so I'll merely point out that the approach I want to advertise is perfectly compatible with imprecise probabilities.

One further trick may be worth dwelling on for a little longer. The idea is to use a two-tiered process in which sensory modules first implement a simplified version of the model I have outlined to estimate the *most probable* hypothesis about the environment in light of the current sensory input. In the second stage, this hypothesis is then treated as the input signal to adjust the subjective probabilities that feed into rational action. Computationally, the two-tiered approach has several advantages. For one thing, the sensory modules can work with simplified, special purpose probability measures that don't have to take into account all the information available to the agent. (We'd effectively return to the "sensor models" from artificial intelligence.) In addition, it is much easier to find a single plausible interpretation of an incoming signal – a single guess about the

10 In a Bayes net, assumptions of conditional independence are directly reflected in the structure of the network, which allows for more perspicuous ways of identifying the input partition in a Jeffrey update. Along these lines, [Schwan and Stern 2017], drawing on [Pearl 1988], suggest that an agent should Jeffrey conditionalize on a partition $\{E_i, \ldots E_n\}$ iff every node A in the network is d-separated from the input node I by the elements E_i of the partition. Schwan and Stern and Pearl, however, do not treat I as an imaginary element of the agent's doxastic space. Rather, they take I to be an ordinary fact about the world of which the agent *could* become certain if only it were represented in her doxastic space. The problem with this model is that there will typically be no such fact I. For example, there is no fact about the world of which my experience of the rain could rationally make me certain.

11 Thanks here to an anonymous referee.

environment – than to calculate to what extent the signal supports every conceivable hypothesis. Given the large amount of data our senses constantly receive, it might make sense for our sensory modules to focus on this simpler, non-probabilistic task. Producing a single guess about the environment might have the further advantage of allowing fast behavioural responses: calculating expected utilities is just as intractable as conditionalization; it is much easier to act on a single hypothesis.[12]

A two-tiered implementation might partly explain why perceptual experiences generally 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 what philosophers mostly have in mind when they talk about perceptual content), but it is clearly not what I conditionalize on, since I do not become certain that one line is longer. I know that the lines are the same length, but the mechanism that produces the categorical interpretation is not sensitive to this knowledge.

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

Here it is important not to conflate different levels of modelling. Hypotheses about the "Bayesian brain" [Doya et al. 2007] are often understood as conjectures about the internal processes involved in perception and action. The perspective I want to defend is largely neutral on these issues. For example, it does not settle whether perceptual input is processed in classical bottom-up style or in the more top-down fashion postulated by recent accounts of predictive coding.

The only suggestion I have for lower-level Bayesian models in cognitive science concerns the interpretation of these models. Cognitive scientists commonly describe perception as a process of inferring facts about the environment from sensory stimuli; they talk about the surprisingness of incoming signals or about the construction of internal models that predict those signals. Taken literally, this suggests that our cognitive system is given direct and infallible information about complicated electrochemical events in its periphery and then faces the task of explaining or predicting the occurrence of these events (see e.g. [Rieke 1999]). Yet most of us are fairly ignorant of the electrochemical processes in our nervous system: the infallible basis of our empirical knowledge appears to get lost through cognitive processing. What I want to suggest is that from the perspective of

12 The intermediate hypothesis produced by the sensory modules need not be a pure, objective proposition about the environment; it might also involve "imaginary" elements, locating the present state of the world in a high-dimensional phenomenal space, as (effectively) suggested e.g. in [Shepard 2001].

a cognitive system, sensory inputs are not represented *as electrochemical events*, even though that is what they are. We can distinguish between the inputs themselves and the corresponding elements in the domain of a system's probability measure. What is "given" in perception are not sophisticated facts about neurophysiology, but imaginary sense-datum propositions that don't directly settle any substantive question about the world.

4 Softcore empiricism

Let me now explore some epistemological consequences of the model I have proposed. As I mentioned in section 1, Jeffrey developed his alternative to classical conditionalization because he rejected the view he called *hardcore empiricism*: that our knowledge of the world rests on a foundation of infallible and indubitable beliefs about present experience. On the most familiar version of the empiricist picture, all our knowledge is derived from an infallible perceptual basis, drawing on a priori connections between the verdicts of experience and statements about the external world. If the 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.

The model I have defended bears a superficial similarity to the classical empiricist picture: experiences confer absolute certainty on a special class of ("imaginary") propositions; beliefs about the external world are then adjusted according to their prior connections with these propositions. But there are also important differences.

First of all, imaginary propositions do not represent true features of the relevant experiences – at least not in a cognitively transparent way. They do not distinguish real ways the world could be at all. (No wonder they can never turn out to be false.) As a corollary, there are no intrinsic logical or analytical connections 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 our prior (conditional) probabilities.

In this way, the model is tailored to accommodate the holism of confirmation. Whether sensory information E supports a genuine hypothesis H about the world always 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 experience are imaginary propositions without real empirical content.

Second, the model I have put forward is not committed to an ontology of sense data or irreducible phenomenal properties. It also 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. Nothing I have said suggests that that we also, or primarily, see non-physical ideas, impressions, or sense-data.

Third, the model I have outlined does not assume that perceivers have special introspective access to imaginary propositions, nor does it assume that such propositions are objects of "belief" in the intellectualist sense that dominates discussions in epistemology. The claim is not that whenever we have a perceptual experience, we become certain of a special "observation sentence" from which we then deduce other, perhaps probabilistic, statements about the world. Conditionalization is not an inference, with premises and conclusion, and it is not supposed to be a conscious, deliberate activity. Perceivers don't need words or concepts that capture the imaginary content of their perceptions, and they don't need to conceptualize their experiences as reasons for their beliefs. As Sellars [1956] and others have pointed out, these commitments render the foundationalist picture highly unappealing.

It is hard to explain the epistemic impact of perceptual experience if we focus on the intellectualist sense of 'belief'. Since perceptions don't seem to have the required sentence-like, "conceptual" content, how can they support or justify beliefs with that kind of content? Is the link between perception and belief merely causal, outside the domain of epistemology? That also seems wrong, for there clearly *are* rational constraints on how one's beliefs may change through perceptual experience. Again, scientists are not free to change their beliefs in any way they please when observing a litmus paper that has turned red.

To make progress on these issues, we should accept that epistemology is not confined to intellectualist belief. The functionalist notion of belief (or credence) popular in the Bayesian tradition brings us one step further, but it still doesn't reach far enough, at least if we restrict it to real-world propositions. As I've argued in sections 1 and 2, how an agent's representation of the world should change under the impact of perceptual experience is not a function of the experience and the agent's representation of the world before the experience. Something else plays a role. In the model I have described, that something else is represented by the agent's extended probabilities.

So what I'm advocating is not a rebranding of classical empiricist foundationalism. In many respects, the model I want to advertise looks more like Jeffrey's "softcore empiricism": it offers a systematic account of how perceptions affect rational attitudes about the world, without making the agent certain of any substantive propositions, and without assuming any fixed probabilistic connection between experiences and propositions about the world.

This brings me back to the problem of priors. Suppose you're a scientist and you've just observed a litmus strip turning red. Absent unusual background assumptions, you should become confident that the strip is red and the tested substance acidic. Why is

that? The model I have put forward does not give an answer. It only says that your new probability in the strip being red should equal your previous probability conditional on the imaginary proposition associated with your experience. But why should that conditional probability be high?

Now, one advantage of the model I have proposed is that the relevant conditional probabilities can themselves be adjusted through learning. So we can explain why the effect of your visual experience is sensitive to background information about the lighting conditions and your eye sight. But that doesn't fully answer the question. Suppose (unrealistically, of course) that your probabilities evolved from an *ultimate prior* probability function by successive conditionalization on sensory evidence. Does this automatically make your probabilities epistemically rational? Arguably not. With sufficiently deviant ultimate priors, your entire history of perception would lead to a state in which you treat the experience of the red litmus paper as strong evidence that the paper is blue (or that Tycho Brahe was poisoned by Johannes Kepler). So there must be substantive, non-formal constraints even on ultimate priors – equivalently, on what one may believe in light of an entire history of sensory input.

How tight are these constraints? Some have argued that there is a unique rational prior, so that rational agents with the same history of sensory input would always arrive at the same credences (e.g. [White 2005]). Others disagree (e.g. [Meacham 2014]). Again, this is not the place to settle these debates. The approach I want to advertise is compatible with either position. I do assume, however, that there are *some* non-formal constraints on ultimate priors.

Where do these constraints come from? It is doubtful that they could be defended by non-circular a priori reasoning. Perhaps they reflect irreducible epistemic norms. Or perhaps they can be explained as (in some sense) constitutive of the relevant intentional states: what makes it true that a given state is a belief that an object is red is in part that it is normally caused by perceptions of red things. We might also try to vindicate some constraints by objective, external correlations. For example, 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 the objective chance of C given S is x, we might say that S objectively supports C to degree x. And so we might say that a subject is justified in assigning conditional credence x to C given $\rho(S)$, absent relevant evidence, iff x matches the degree to which S objectively supports C.[13]

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

13 See [Dunn 2015], [Tang 2016], and [Pettigrew Ms] for further development and exploration of this idea.

with 21st century cognitive science.

5 Puzzles of consciousness

Before concluding, I want to explore one further application of our model, to the puzzle of conscious experience.

I have suggested that terms like 'credence' or 'belief' might be reserved for an agent's attitudes towards real propositions, since the non-real part of an agent's probability function does not serve a straightforwardly representational function. However, this difference between real and imaginary propositions need not be transparent to the agent herself. An agent's cognitive system need not draw a sharp distinction between the two kinds of attitudes. Our robot, for example, does not need a special database for real propositions in addition to its database for imaginary (and complex) propositions. From the robot's perspective, it might simply appear as if reality had an extra dimension, an extra respect of similarity and difference. Perceptual experience will then appear to convey direct and certain information about this aspect of reality, and only uncertain information about everything else. Conversely, ordinary information about the world will never suffice to fix the apparent further dimension of reality: there is no conjunction of real propositions conditional on which any sense datum proposition has probability 1. The robot will therefore be tempted to conclude that physics is incomplete: that there are special phenomenal facts revealed through experience that are not implied by or reducible to physical facts about the arrangement and dynamics of matter. Yet our robot could well exist in a completely physical world.

Perhaps we all are in a position not unlike this robot. Our perceptual experiences do appear to convey a special kind of information that is more certain than our ordinary beliefs about the world. To illustrate, consider your present perceptual experience. Are there any possibilities you can conclusively 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 systematic hallucinations, rewired brains, evil demons or the like. It could be a real situation from the past, if you ever went skiing. Your experiences in that 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, it seems that we can rule out any situations in which we have a sufficiently different experience. 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] put it: "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.

The model I have put forward suggests a different response. Both the reading experience and the skiing experience are associated with imaginary propositions in your extended doxastic space, namely, the propositions of which either experience would make you certain. When you entertain the hypothesis that you are in the skiing scenario, what you entertain includes the relevant imaginary propositions. And these propositions are incompatible with the imaginary propositions associated with the reading experience: their conjunction is the empty proposition. Hence the reading experience allows you to conclusively rule out the skiing scenario – not by its physical features, but by its "imaginary features", so to speak.

Along the same lines we can explain other phenomena that seem to put pressure on physicalism. Consider Mary (from [Jackson 1982]), who has learned all physical facts about colours and colour vision without having seen colours. If Mary's probability space has an imaginary dimension, her physical knowledge may still leave open many possibilities along the imaginary dimension. For example, let I_R be an imaginary proposition associated in Mary's cognitive system with experiences of physical type X, typically caused by looking at red things. In Mary's extended probability space, this association will be contingent: $P(I_R/X) \ll 1$. So when she is eventually presented with some coloured chips, without being told their colour, she will become certain of I_R, but she won't be able to tell that she is in physical state X. Conversely, if she learns that she will be in state X tomorrow, this will leave her uncertain about I_R-*tomorrow*. Her apparent ignorance will only be resolved when she looks at something of which she knows the colour. All this will be so even if Mary lives in a completely physical world.

Similarly, if I_R is an imaginary proposition associated with red experiences, and P is the totality of all physical truths, we can explain why both $P \,\&\, I_R$ and $P \,\&\, \neg I_R$ are a priori conceivable (see [Chalmers 2009]), even if the world is completely physical.

In sum, the phenomena that appear to support dualism about consciousness might be artefacts of the way we process sensory information.

To be clear, the model I have outlined makes no direct claims about consciousness. I never mentioned consciousness when I introduced the model. Moreover, empirical evidence (e.g. about binocular rivalry, see [Blake and Logothetis 2002]) makes clear that our conscious experience does not simply track the stimulation patterns in our

sense organs. Consciousness rather seems to play a role in something like the two-stage processing about which I speculated in section 3. There I suggested that our sensory systems might compute a single (course-grained) hypothesis about the environment which is then turned into an input signal for personal-level probabilities as well as possibly driving immediate behavioural reactions. Clearly, this is mere speculation. I have no qualified views on the functional role of consciousness in our cognitive architecture, and I don't claim that my model makes much progress on this issue – on what Chalmers [1995] calls the "easy problem" of consciousness. But it might help with the "hard problem", the problem of explaining how physical processes in the brain seem to create a phenomenon that can't be understood in physical or functional terms at all. My suggestion is that, for reasons to do with the efficient processing of sensory signals, our subjective picture of the world has an added dimension that makes it appear as if perceptual experiences carry a special kind of information that goes beyond physical and functional information.

I want to close with one more puzzle about consciousness that has not received much attention.[14] 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 experience with your present reading experience. Both experiences have a distinctive phenomenal character. For the skiing experience, 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 cause the experience and with 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.

14 [Latham 2000] discusses a version of the puzzle.

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 what it's worth, I have asked a number of philosophers who prominently reject the a priori entailment of the phenomenal by the physical, and they all agreed that their view makes the scenario I described epistemically possible.)

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 listen to the news in the morning, they have the experience of still sleeping; when they have started working, they have the experience of eating breakfast and listening to the news, and so on. How convenient that we don't live in a world like that! (If indeed we don't. 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 conjoined with a reading scenario. 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 and thus on the probabilities of real propositions conditional on imaginary propositions. (Recall the discussion of non-formal constraints in section 4.) Absent unusual background information, a specific sense datum must be regarded as strong evidence for a narrow range of hypotheses about the world. 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 – individuated by the associated imaginary propositions – is caused by an unusual environment. The inverted qualia worlds just described are extreme scenarios of this type. The model I have outlined suggests that they must have negligible probability. They are *almost* a priori ruled out.

6 Conclusion

Much of what we know about the world we know through perception. But how does perception provide us with that knowledge? Perceptual experiences do not seem to deliver direct and certain information about the external world. The old empiricists held that perceptual experience instead delivers information about an internal world of sense data, from which we infer hypotheses about the external world. But the idea of a luminous internal world is hard to square with a naturalistic picture of cognition and with our general fallibility. A more sober response rejects the assumption that rational belief requires a bedrock of certainty: perceptions may increase or decrease the credibility of various external-world hypotheses without rendering anything certain. Unfortunately, once we try to fill in the details in this response, we run into serious problems. On closer inspection, an agent's belief state prior to a given experience does not contain enough information to settle how the beliefs should change through the experience.

I have proposed a way out that takes a step back towards the old empiricist account. In order to adequately respond to sensory stimulation, I have argued, it can be useful

to extend the domain of an agent's subjective probability function by an "imaginary" dimension whose points are associated with sensory signals in such a way that when a signal S arrives, a corresponding "imaginary proposition" $\rho(S)$ becomes certain; the probability of real propositions is then adjusted in accordance with their prior probability conditional on $\rho(S)$. The space of imaginary propositions plays an epistemological role somewhat analogous to the empiricist's internal world. In section 5, I suggested that the similarities might go even further insofar as imaginary propositions might correspond to the phenomenal properties that appear to present themselves to us in perception. In other respects, however, the picture I have outlined is closer to Jeffrey's "softcore empiricism" than to traditional flavours of empiricism, as I've argued in section 4.

The model I have presented is abstract and formal. It does not settle that – let alone explain why – a given type of red experience should make an agent confident that she confronts something red. Nor does it imply any particular algorithm or mechanism for computing the new probabilities. As such, it remains neutral on many contentious and difficult questions in epistemology and cognitive science. Nonetheless, it may provide a useful framework for tackling some of the more substantive questions in these areas.

References

Randolph Blake and Nikos K. Logothetis [2002]: "Visual Competition". *Nature Reviews Neuroscience*, 3(1): 13–21

David Braddon-Mitchell and Frank Jackson [1996]: *Philosophy of Mind and Cognition*. Oxford: Blackwell

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

— [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

Andy Clark [2013]: "Whatever next? Predictive brains, situated agents, and the future of cognitive science". *Behaviour and Brain Science*, 36: 181–204

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

Kenji Doya, Shin Ishii, Alexandre Pouget and Rajesh P.N. Rao (Eds.) [2007]: *The Bayesian Brain*. Cambridge, Mass.: MIT Press

Jeff Dunn [2015]: "Reliability for degrees of belief". *Philosophical Studies*, 172(7): 1929–1952

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

Thomas L. Griffiths, Charles Kemp and Joshua B. Tenenbaum [2008]: "Bayesian models of cognition". In R. Sun (Ed.) *Cambridge handbook of computational cognitive modeling,* Cambridge University Press, 59–100

Thomas L Griffiths, Falk Lieder and Noah D Goodman [2015]: "Rational use of cognitive resources: Levels of analysis between the computational and the algorithmic". *Topics in cognitive science*, 7(2): 217–229

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

Jakob Howhy [2014]: *The Predictive Mind*. Oxford: Oxford University Press

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

James Joyce [2005]: "How probabilities reflect evidence". *Philosophical Perspectives*, 19: 153–178

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

David Lewis [1974]: "Radical Interpretation". *Synthese*, 23: 331–344

— [1979]: "Attitudes *De Dicto* and *De Se*". *The Philosophical Review*, 88: 513–543

— [1983]: "New Work for a Theory of Universals". *Australasian Journal of Philosophy*, 61: 343–377. Reprinted in [Lewis 1999]

— [1995]: "Should a Materialist Believe in Qualia?" *Australasian Journal of Philosophy*, 73: 140–144

— [1999]: *Papers in Metaphysics and Epistemology*. Cambridge: Cambridge University Press

Christopher Meacham [2014]: "Impermissive Bayesianism". *Erkenntnis*, 79: 1185–1217

Judea Pearl [1988]: *Probabilistic reasoning in intelligent systems: Networks of plausible inference*. San Mateo: Morgan Kaufmann

Judea Pearl and Stuart Russell [2001]: "Bayesian Networks". In M. Arbib (Ed.) *Handbook of Brain Theory and Neural Networks,* Cambridge (MA): MIT Press

Richard Pettigrew [Ms]: "What is justified credence?" Manuscript

Fred Rieke [1999]: *Spikes: Exploring the neural code*. Cambridge (MA): MIT press

Stuart J. Russell and Peter Norvig [2010]: *Artificial Intelligence: A Modern Approach*. Cambridge (MA): MIT Press, 3rd edition

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

Ben Schwan and Reuben Stern [2017]: "A Causal Understanding of When and When Not to Jeffrey Conditionalize". *Philosopher's Imprint*, 17(8)

Wolfgang Schwarz [2017]: "Diachronic norms for self-locating beliefs". *Ergo*, 4

Matthias W. Seeger and David P. Wipf [2010]: "Variational Bayesian inference techniques". *Signal Processing Magazine, IEEE*, 27(6): 81–91

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

Roger Shepard [2001]: "Perceptual-cognitive universals as reflections of the world". *Behavioral and brain sciences*, 24(4): 581–601

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

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

Scott Sturgeon [2008]: "Reason and the grain of belief". *Noûs*, 42(1): 139–165

Weng Hong Tang [2016]: "Reliability theories of justified credence". *Mind*, 125(497): 63–94

Ed Vul, George Alvarez, Joshua B Tenenbaum and Michael J Black [2009]: "Explaining human multiple object tracking as resource-constrained approximate inference in a dynamic probabilistic model". In *Advances in neural information processing systems,* 1955–1963

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

Roger White [2005]: "Epistemic Permissiveness". *Philosophical Perspectives*, 19: 445–459

Alan Yuille and Daniel Kersten [2006]: "Vision as Bayesian inference: Analysis by synthesis?" *Trends in Cognitive Sciences*, 10(7): 301–308

Hao Zhang, Hongzhe Dai, Michael Beer and Wei Wang [2013]: "Structural reliability analysis on the basis of small samples: an interval quasi-Monte Carlo method". *Mechanical Systems and Signal Processing*, 37(1-2): 137–151