Imaginary Foundations

Wolfgang Schwarz*
Draft, 10/06/2013

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? After all, many other types of liquid look just like water. It could be a stream of vodka. Of course that is an outlandish possibility. Nevertheless, if for whatever reason I had taken it seriously before looking outside, my visual experience wouldn't put me in a position to rule it out. So if we understand 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. The windows could have been replaced with sophisticated LCD screens. My claim is not that these are possibilities I should take seriously, or that I don't *see* or *know* that it is raining. All I claim is that my visual experience by itself doesn't put me in a position to rule out these possibilities. What, then, is the information provided to me by my senses?

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 (what if I'm dreaming?), 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.

But this view faces serious 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?

So perhaps we started with a bad question. Perhaps my visual experience doesn't put me in a position to exclude possibilities at all – not conclusively, and not by itself. Given my background beliefs, my experience allows me to conclude, tentatively and defeasibly, that it is still raining. In general, experiences combine with background beliefs to confer certain degrees of plausibility or probability to various claims about the world.

* Acknowledgements.

The framework of Bayesian epistemology allows us to put these issues into sharper focus. Assume beliefs come in degrees, measured by numbers between 0 and 1. Assume further 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 that 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*. Here E is meant to capture the information delivered by the experience. Since its new probability is 1, E can't be ordinary information about the world. When I looked outside the window, I didn't become absolutely certain that it is raining water. (Indeed, my experience *raised* the probability of some alternative hypotheses such as 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 the objects of conditionalisation.

To avoid this consequence, Richard Jeffrey proposed an alternative to classical Bayesianism which he called *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.

Often an experience is directly relevant to some propositions and not to others. Your visual experience of the table cloth reveals something about the cloth's colour, but nothing about whether there is water on Mars. 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 are 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, this last name is rather misleading.[1]

1 For ease of exposition, I have assumed that the experience is directly relevant only to a finite partition

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.

However, the new probabilities $x_1, \ldots, x_n$ for $E_1, \ldots, E_n$ clearly 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 whether the owner of the room generally likes green tablecloths. So how does a given 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. It was already raised by Carnap in his 1957 correspondence with Jeffrey (published in [Jeffrey 1975]), where Carnap reports that he wasn't happy with the classical Bayesian picture but felt compelled to stick with it because he couldn't find an answer to the input problem. Since then, nobody else has found a plausible answer either, and it is widely thought that the problem simply can't be solved.

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. Intuitively, the reason is that the second experience is in all relevant respects just like the first and thus provides little new information. But this fact about the two experiences need not be entailed by 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.[2]

$E_1, \ldots, E_n$, but the formula is easily extended to infinite cases. Countably infinite partitions raise no special problems; for uncountable partitions, orthodox probability theory requires that the associated new probabilities are given by a probability distribution over a sigma-algebra generated by the partition (a sub-sigma-algebra of the agent's complete doxastic space). The sum in Jeffrey's rule is then be replaced by a Lesbesgue integral over this algebra; see [Diaconis and Zabell 1982: sec. 6].

2 The present considerations are inspired by [Garber 1980], which shows 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

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. Among other things, there are norms on how one's beliefs may 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 supported 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?" – "Oh, I looked at the rain".)

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.

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

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]. Perhaps the most serious 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).

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 produce different patterns of light waves activating the photoreceptors, which in turn lead to a different combined signal produced by the sense organ.

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 shouldn't become absolutely certain that a particular one of $E_1, \ldots, E_n$ obtains, giving zero probability to all the others. Instead, we could implement a form of Jeffrey conditionalisation, where the new probabilities $x_1, \ldots, x_n$ over $E_1, \ldots, E_n$ reflect the expected relative frequency (or, if you want, objective chance) with which the conditions obtain when the signal is produced.[3]

But this is 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 $x_1 = x_2$. Nonetheless, before the arrival of S the robot may have received information that supports E_1 over E_2. On the present account, the new signal will wipe out this information, setting $P_{new}(E_1) = P_{new}(E_2)$. Relatedly, our

3 The list of alternatives $E_1, \ldots, E_n$ may not be exhaustive, as 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 probably can't guarantee that S couldn't also be triggered "erroneously", say if the robot is hit with a bat while running out of battery. So we should set $x_{n+1} > 0$ and adjust the other probabilities accordingly.

robot should be able 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. Initially, before the robot has encountered further information, these conditional probabilities for E_i given S 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. If on a given occasion the robot has other evidence favouring E_2 over E_1, its conditional probability for E_2 given S may, on this occasion, be higher than the probability for E_1 given S. Receiving S then won't overwrite the earlier evidence. In principle, the update could still take the form of Jeffrey conditionalisation, but the $x_1, \ldots, x_n$ values are now given by $P_{old}(E_1/S), \ldots, P_{old}(E_n/S)$. Yet 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, $P_{new}(A) = P_{old}(A/S)$.

As it stands, this proposal has an odd consequence. It requires our robot to have sophisticated opinions about the electrical signals S produced by its sense organs. However, a little reflection reveals that this is not actually needed. Suppose the robot's probabilities were originally defined only for certain propositions about the environment. The present proposal requires that we extend this space by new propositions corresponding to the robot's sensory signals so that whenever signal S occurs, the robot conditionalises on the corresponding proposition. This *causal* correspondence between the signals and the added propositions is all we need. It is not required that the new propositions intrinsically describe or identify the relevant electro-chemical signals. In fact, they don't need to be genuine propositions at all: they don't need to say anything about the world.

Think of it this way. The robot's probability function is supposed to capture its probabilistic views about the world. So it should assign probabilities to genuine 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 various combinations of these elements with the original propositions 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 $\neg A \,\&\, E_S$, and E_S is the new element associated with signal S (i.e., the element whose probability is raised to 1 by signal S), then receiving S will make it 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.

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 **R** is closed under Boolean operators, so we can identify the propositions in **R** with sets of "possible worlds" (the atoms of the algebra). Let **E** be a set of arbitrary objects (disjoint from **R**) that stand in a one-one correspondence to the signals the agent can receive. Each pair $\langle w, e \rangle$ of a possible world w with an element e of **E** is 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 **R**) are *real propositions*; subsets of **E** are *imaginary propositions*; members of **E** are *sense data*. In an agent's complex doxastic space, a real propositions A can be 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.[4]

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. 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 and will not be important for the present topic. (See [Schwarz 2012] for my own proposal and some alternatives.)

This completes my proposal. This is how I would design an ideal agent. I can't take credit for the invention, however. The ideas I have outlined are well-known in artificial intelligence (see e.g. [Russell and Norvig 2004], esp. ch.15).[5] Given their advantages and

4 The construction of complex propositions is analogous to a common construction of *centred* propositions as sets of pairs of a possible word w and a centre c; here, too, one can usefully identify the uncentred proposition A with the set of centred worlds whose world coordinate lies in A.

5 Computer scientists usually take the agent's conditional probabilities $P(A/S)$ over real propositions A given a signal S as derived from inverse probabilities $P(S/A)$ of signals given hypotheses about the world; the update then goes via Bayes' Theorem: $P'(A) = P(A/S) = (P(S/A)P(A))/(\sum_i P(S/A_i)P(A_i))$. Since the posterior $P'(A)$ is often written simply as $P(A/S)$, the assumption that probabilities change by conditionalisation is here implicit in the notation.

the importance of perception to survival, one might expect that nature has implemented similar mechanisms in ourselves. Research in cognitive science appears to bear out this expectation. Here our model is known as *the Bayesian model of perception* and has proved successful in explaining and unifying many aspects of the sensory processing in our brain (see e.g. [Kersten et al. 2004] for a survey).[6]

3 Interpretation and implementation

I have outlined a model of how subjective probabilities change under the impact of sensory stimulations. The model is in the first place a normative ideal: a design template for an ideal computational system. But it is also a plausible, albeit no doubt imperfect, 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, terms like 'belief' and 'credence' are often used in a demanding sense tied to conscious thought and linguistic assertion. On this usage, the objects of credence are often identified with sentences in the agent's public or private language; to say that the agent assigns high credence to a sentence might then be taken to mean that she feels a high degree of conviction when contemplating the sentence or that she is inclined to assert that the sentence has high probability. The probabilities in the Bayesian model are clearly 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 a theoretical term characterised by it role in a comprehensive psychological theory concerning an agent's subjective representation of the world, her goals, her perceptual inputs and behavioural outputs. On this account, belief is not 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 as propositions, i.e. ways things could be. To assign a subjective probability to a proposition, you don't need to form an explicit judgment, nor do you have to somehow "grasp" the proposition. All that's required is that your cognitive state plays the right causal-functional role (on this point see also [Stalnaker 1984: ch.1]).[7]

6 A caveat: in cognitive science, the postulated probability function that gets conditionalised on sense data is usually not identified with the subject's credence function. I think the scientific model is *compatible* with this interpretation, but for reasons discussed in the next section the issue is not completely straightforward.

7 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 suggests that this is not true for the probabilities involved in perception; see [Knill and Pouget 2004].

The Bayesian model of perception suggests that there are reasons to extend an agent's probabilities beyond real propositions. We could still reserve 'degree of belief' and 'credence' for 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 which 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 part of it 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.[8] 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 bear on propositions about the world. In the Bayesian model, the agent can change her mind about how likely she deems this or that sense datum given different a hypothesis about the world.

It is reasonable 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 doesn't 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 simply has 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

8 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. Assuming that the agent's Ramsey-style probabilities are defined over a complex space with an added imaginary dimension, the agent might also have words or "concepts" to reason about complex and imaginary propositions. On the other hand, her language may well 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 an abstract, structural level. It doesn't say which real propositions are normally supported by which perceptual experience. It doesn't say whether those real propositions are qualitative or singular. More importantly, it doesn't imply that perceptual experiences are the only force that acts on subjective probabilities. For example, a different kind of update is plausibly called for in practical deliberation (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, it might furthermore 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. To overcome these problems, we could start by breaking up, distributing and simplifying the central database. Instead of encoding precise probabilities for every proposition, it might be enough to encode some basic constraints on probabilities: that p is more probable than q, that r is probabilistically independent from s, and so on.[9] To aid the update on sense data, we could use a designated 'sensor module' that only

9 The agent's opinions about the world will then be captured by a *set* of probability functions: all the functions that meet the constraints; see [Jeffrey 1984].

makes use of conditional probabilities for sense data given a restricted set of (perhaps somewhat unspecific) hypotheses about the immediate present environment, so that in a first step only the probability for these hypotheses get conditioned on the sensory signal. Here it will also help to represent sense data not just as bare tags, but in a reasonably structured way, mirroring structural aspects of the signals. Note also that the signals in the Bayesian model don't have to be identified with excitation patterns of receptor cells. We could add a hard-coded pre-processing step before the probabilities come into play; the sense data on which the agent conditionalises would then match the output of the pre-processing step; this output would play the role of the signal in the Bayesian model.

Tricks like these will reduce the computational costs, but for some occasions it might be good to have mechanisms that are still cheaper and faster. If it looks like a rock is about to hit your head, your top 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. Computationally, it is much easier to find a single plausible interpretation of the signal – a single guess about what type of environmental conditions may have produced the signal – than to figure out to what extent the signal supports each such hypotheses over the alternatives. It is also much easier to act on a single proposition about the environment than on a probability distribution: calculating expected utilities is computationally just as hard as conditioning. For agents like us and our ancestors for whom it is often vital to respond very quickly to incoming signals, it will be desirable to have such a fast and simple "dogmatic" mechanism on board, alongside the slower probabilistic architecture. This might explain why many perceptual experiences intuitively come with a built-in interpretation about the environment: 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 kind of "perceptual content" is clearly not what I conditionalise on. 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 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 it might account for salient features of our psychology that are not predicted by the simple Bayesian ideal.[10]

10 For the most part, the tricks I mentioned are familiar in cognitive science and computer science. The simple Bayesian ideal can still play an important explanatory role here, since many peculiarities of our cognitive architecture can be explained as artifacts of how our brain approximates the ideal (see [Gershman and Daw 2012]). For example, an efficient way to approximately compute conditional probabilities is Monte Carlo simulation; [Gershman et al. 2012] argue that if our perceptual system employs this method in the processing of ambiguous figures like the duck rabbit, then one interpretation of the figure should usually dominate, although which one dominates occasionally changes without any change in the perceptual input – which is precisely what we find (see also [Sanborn et al. 2010]).

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 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 are no logical or semantic 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 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 themselves were 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 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. If you are convinced that there is an odd number of stars in the universe merely because your ultimate prior already assigned high probability to this proposition then your conviction isn't justified. As I mentioned in section 1, there are substantive epistemic norms on rational credence that constrain what one may believe in response to given observations. Often such norms concern the bridge principles linking sensory stimuli to propositions about the world. Absent unusual background beliefs, a visual experience of rain should not make one confident that Tycho Brahe was poisoned by Johannes Kepler. In the Bayesian model, these norms map onto constraints on prior probabilities conditional on imaginary propositions.

Where do these constraints come from? It is doubtful that they could be defended by non-circular a priori reasoning that would convince a skeptic, 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. My own view is that some such norms have to be taken as basic.

Alternatively, we might try to vindicate at least some of them 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 [**?**: xxx] 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.[11]

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 have the required sentence-like, "conceptual" content, it looks like they can't support or justify beliefs; the link between perception and belief must 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 disappears. One can model perceptions as content-bearing states without implying that they are somehow directed at 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 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, even if we don't count attitudes towards imaginary propositions as 'belief'.

5 Puzzles of consciousness

Is there any introspectible trace of our alleged certainty in imaginary propositions? Arguably yes. Consider your present perceptual experience. Are there any possibilities

11 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.

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 given your experience. For example, consider a scenario in which you are skiing – a normal skiing scenario, without pervasive halluzinations, 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 intuitively rule out any situations in which we have a suitably different experience. That's 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.[12]

Our Bayesian model shows that physicalism is in fact compatible with 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 epistemic 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 epistemic space are ruled out not by their physical features, but by their imaginary features.

These considerations suggest that the Bayesian model might also explain other well-known facts about our epistemic situation that are often said to pose a threat to physicalism. Imagine an agent with the kind of architecture I have outlined, in a completely physical world. To this agent, it will seem as if there are facts that go beyond ordinary, physical facts about the 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 epistemic space. Bayesian Mary could have known all

12 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 [**?**]).

physical facts about red experiences and still have learned something new when she first sees colours ([Jackson 1982]). Similarly, if Q is an imaginary proposition associated with red experiences, and P is the totality of all physical truths, both $P \,\&\, Q$ and $P \,\&\, \neg Q$ will be a priori conceivable for Bayesian agents ([Chalmers 2009]).

On the other hand, which imaginary proposition a Bayesian agent learns on a given occasion is completely determined by her biological and physical features at that time; in the simplest case, by her sensory stimuli. Hence the agent might discover that the non-physical dimension of possibilities concerning her experiences are invariably tied to the presence of certain physical features. She might also note that the realm of the physical appears causally closed in a way that makes it look like non-physical properties would have to be causally idle. She might thus be tempted to hypothesise that the apparently non-physical properties revealed through experience are after all identical to the physical properties with which they are invariably associated. But she might also note that this identity hypothesis does not fit the usual pattern of a posteriori identities. In reality, the non-physical properties that figure in the agent's epistemic space are an artifact of the way she processes sensory signals.

A subtle question is whether, on this picture, the intuitive judgments that appear to support dualism are mistaken. It depends on whether we evaluate these judgments relative to the agent's extended epistemic 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. The intuition that Mary is still ignorant about any real, objective facts once she knows all the physical facts is mistaken. On the other hand, relative to Mary's own epistemic space, it is perfectly true that Mary learns something new when she first sees colours.

Imaginary information is in many ways analogous to indexical information.[13] Arguably, our epistemic space includes not only propositions about the universe as a whole, but also propositions about our own location in space and time. If you know every truth about the universe from a God's eye perspective, you may still not know who in the world you are and what time is now. When you find out these things, 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 epistemic space, finding out what time it is excludes genuine possibilities – albeit not objective possibilities about the world.

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

13 Recall footnote 4 above.

"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 conditioning 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]

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. Perhaps that role is 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 [**?**], the present model only solves the "hard problem" of consciousness, not the "easy problem".

I want to close with a few remarks on another puzzle about consciousness that has not received much attention in the literature (none, as far as I know). The puzzle is the apparent *fit* between the phenomenal features of mental states and their functional role. To see what I mean, compare again the skiing scenario with your actual present situation. 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] It is then just a coincidence,

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 strategy is to explain how this opaque reference to physical states is supposed to work (see [Chalmers 2007], [Levine 2007]). In my view, it is best to drop that claim about reference.

15 Strictly speaking, one could deny that the phenomenal is entailed by the physical but also deny the

epistemically speaking, 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. 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. For can we really be sure? Wouldn't our phenomenal experiences be just as they actually are?)

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, there are 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 perhaps they can't be completely rewired. In any case, 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 – much more so than scenarios involving evil demons or brains in vats – and thus deserve only negligible a priori probability. This is why the scenarios are so odd. They are *almost* a priori ruled out.

References

David Chalmers [2007]: "Phenomenal Concepts and the Explanatory Gap". In *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism,* Oxford University Press

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

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.

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

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

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

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

— [1984]: “Bayesianism with a Human Face”. In J. Earman (Ed.) *Testing Scientific Theories,* Minneapolis: University of Minnesota Press, 133–156. Reprinted in [Jeffrey 1992]

— [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

Richard C. Jeffrey [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

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

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”. London: Routledge & P. Kegan

S.J. Russell and P. 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

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

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

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