Review of Philosophy and Psychology
https://doi.org/10.1007/s13164-022-00647-9

Representing Probability in Perception and Experience

Geoffrey Lee[1] · Nico Orlandi[2]

Accepted: 23 May 2022

Abstract

It is increasingly common in cognitive science and philosophy of perception to regard perceptual processing as a probabilistic engine, taking into account uncertainty in computing representations of the distal environment. Models of this kind often postulate probabilistic representations, or what we will call probabilistic states,. These are states that in some sense mark or represent information about the probabilities of distal conditions. It has also been argued that perceptual experience itself in some sense represents uncertainty (Morrison *Analytic Philosophy* 57 (1): 15 48, 2016). In this article, we will first consider three models of sensory activity from perceptual neuroscience, namely signal detection theory (SDT), probabilistic population codes (PPC), and sampling. We will then reflect on the sense in which the probabilistic states introduced in these models are probabilistic representations. To sharpen this discussion, we will compare and contrast these probabilistic states to credences as they are understood in epistemology. We will suggest that probabilistic representation, in an appropriately robust sense, can be understood as a form of analog representation. In the last part of the paper, we apply this to the issue of whether conscious experience represents uncertainty—we will interpret this as the claim that there are phenomenal features of experience that serve as analog probabilistic representations.

It is increasingly common in cognitive science and philosophy of perception to regard perceptual processing as a probabilistic engine, taking into account uncertainty in computing representations of the distal environment. Models of this kind often postulate *probabilistic representations*, or what we will call *probabilistic*

Authors are listed in alphabetical order. This article is fully collaborative.

✉ Nico Orlandi
norlandi@ucsc.edu

Geoffrey Lee
geoffrey_lee@berkeley.edu

1 UC-Berkeley, Berkeley, CA, USA

2 UC-Santa Cruz, Santa Cruz, CA, USA

Published online: 29 September 2022

states. These are states that in some sense represent information about the probabilities of distal conditions. It has also been argued that perceptual experience itself in some sense represents uncertainty (Morrison 2016).

Our concern in this paper is to clarify what these representational states would consist in, were they to exist. While the notion of *representation* is notoriously problematic, the theoretical options for understanding it are familiar and widely discussed. Much less familiar, on the other hand, is how to understand the idea that perceptual, sensory and conscious states are *probabilistic* representations. What would it mean for the perceptual system to hold to a probability of 0.6 that there is a vertical line in the visual field? What would it be for a perceiver to consciously perceive a vertical line in a way that involves uncertainty?

Our proposal is that probabilistic representations can be understood as a type of *analog representation*, which *systematically* (through structural resemblance) and *exclusively* represent the probabilities of relevant hypotheses. An example might be a neural firing rate whose strength systematically indicates the probability that a feature is present, and whose function is exclusively to indicate this probability (see Sections 4 and 5). To count as a probabilistic representation in the relevant sense, this type of analog state must also be appropriately *exploited* by the system, in the sense that the state is used in a way that appropriately reflects the probabilistic information that the state carries. In line with this proposal, we also argue that *conscious representation of uncertainty* is best understood in terms of the claim that there are phenomenal features of experiences that serve as analog probabilistic representations.

Importantly, we will *not* defend the existence of probabilistic representations in either subpersonal processing or in conscious experience which is a matter of controversy. Our goal is to clarify what would even count as a probabilistic representation, thereby clarifying the target of the debate. That said, some of the reasons why theorists have postulated probabilistic representations will be discussed when they are relevant.

The first part of the paper considers the probabilistic states postulated by three prominent computational models of perceptual processing while the second part concerns conscious probabilistic states. We will refer to the former states as "subpersonal" but nothing here turns on how this concept is understood; roughly, these states are typically unconscious and inaccessible to the subject, and in that way contrast with the conscious states we consider in the second part. Importantly, we think that our account of the representational status of states postulated by these computational models transfers over in a natural way to the case of conscious representations—this is the motivation for treating both cases in a single paper. If conscious states are simply states of perceptual processing that meet certain functional conditions (e.g. they are globally broadcast), then this kind of unity is exactly what one would expect. As we will discuss though, the additional functional constraints required for consciousness might also turn out to restrict the kinds of representations that can be conscious—in particular, it could be that they require consciousness to give a single definitive interpretation of the world.[1]

[1] For some of the difficulties in making sense of the distinction between personal and subpersonal states see Colombo 2013 and Drayson 2012.

In the current empirical literature there are three prominent models that explicitly propose a view of how probabilistic information could be encoded in subpersonal perceptual processing: signal detection theory (SDT), probabilistic population codes (PPC) and sampling models. Because these models are not presented with much clarity in the philosophical literature, we will describe them in some detail. To be clear, our interest is not primarily in the different possible algorithmic strategies for probabilistic *computation*, but rather the different possibilities for representing probability itself: what do the probabilistic states in these models consist in? That said, there is an intimate relationship between the individuation of a representation and how it is used computationally—for this reason, we will need to discuss relevant computational details. Relatedly, there are some well-known probabilistic computational models (such as predictive processing models, which are covered in other articles in this special issue,) which do not contain a distinctive proposal for how probability is represented, and so we do not consider them here (but see Orlandi and Lee 2018).[2]

Once we have candidate examples of probabilistic states on the table, we will consider in more detail the question of what makes them probabilistic representations. We will approach this issue from two angles. On the one hand, an obvious point of reference is the extensive discussion of the individuation of credences, or *degrees of belief*, in the formal epistemology literature, which are after all a kind of probabilistic mental state. Can sub-personal probabilistic states be treated in a similar way? The other angle we will take is via general theories of what subpersonal perceptual representations consist in—we think they can be adapted in reasonably straightforward ways to allow for probabilistic representation. As mentioned, our proposal is that probabilistic representations can be understood as a form of *analog representation,* and we will take it to be distinctive of analog representations that they are states that *systematically* – through *structural resemblance* (Beck 2019) – and *exclusively* carry information about what they represent. We will elaborate these conditions in Sections 4 and 5.

Before we proceed, two points of clarification. First, our main focus here will be on *occurrent events* that play a representational role, such as patterns of neural firing representing the probability that a certain feature is present, as opposed to the dispositional or implicit kind of representation exemplified by neuronal connection strengths, which might be held to represent conditional probabilities (see Icard 2016). We will briefly comment on this latter case below (p. 13). We are tempted to call the representational states we examine "explicit" probabilistic representations. However, we avoid using this term because in some recent empirical discussions (e.g. Ma 2012) the notion of an "explicit" probabilistic representation is reserved for cases where a statistic such as the mean or variance has been explicitly computed from a probability distribution. In this narrow sense, our probabilistic-states would *not* be explicit.

This leads to our second point of clarification. To keep things simple, we will mostly talk as if the probabilistic states we are discussing are all representations of

[2] There is also some literature on whether predictive processing models posit representations at all. See Gładziejewski (2016), Kiefer and Hohwy (2018) and Kirchhoff and Robertson (2018).

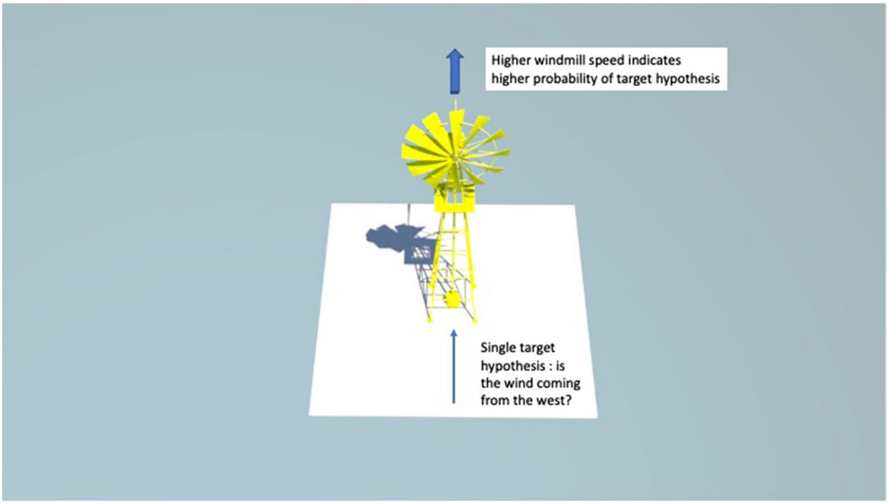

Fig. 1 Windmill used as signal detection system

posterior probabilities (e.g. the probability that a certain distal element like a vertical line is present). However, theories on which representations of likelihood functions, and higher-order statistics such as the mean or variance of a distribution, are also common in the kind of modeling we are interested in. We intend our account to apply to these cases as well (so when we say "probabilities", we often really mean "probabilities or likelihoods").[3]

1 The Varieties of Subpersonal Probabilistic States

We begin by describing three theories: signal detection theory, probabilistic population codes and sampling models. These are computational models designed to help explain how, in the presence of underdetermination and noise, neurons in sensory areas of the brain represent information in a form that allows the reliable computation of an accurate perception of distal stimulus features, and allows the perceiver to successfully make decisions and judgments based on this perception.

To explain in an intuitive way the mechanics of the three models, and to make vivid why they are the natural options to consider, we will use an extended metaphor

[3] We anticipate some scepticism about whether likelihoods are the kind of feature that could be presented in experience (e.g. Morrison (personal communication) suggested this scepticism to us). If the likelihood is understood as the probability of our sensory evidence given the hypothesized state of the world, then one might think that representing it in experience would require a strange kind of self-referential content. Whereas, posterior probabilities can just be more straightforwardly attributed to external events (it might be thought). However, in this paper we reject the view that some such analysis of what a likelihood is (or what a posterior is, for that matter) would have to be explicitly reflected in phenomenology. We return to this point in Section 5.

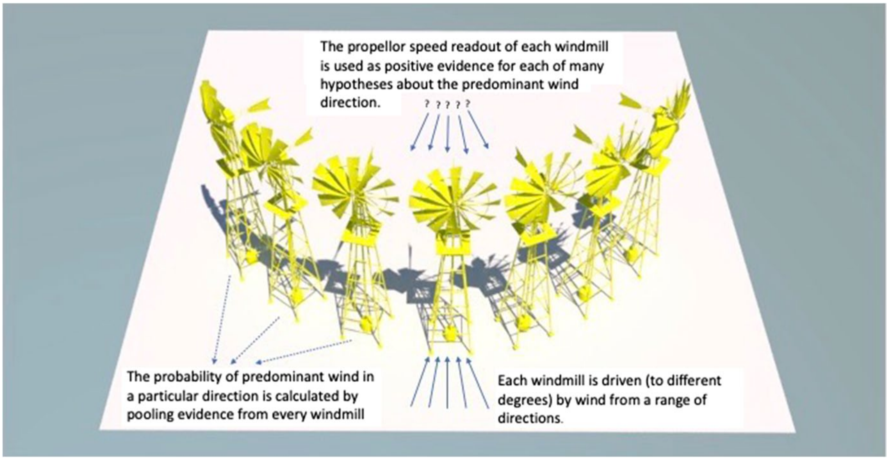

Fig. 2 Windmills used as a probabilistic population code

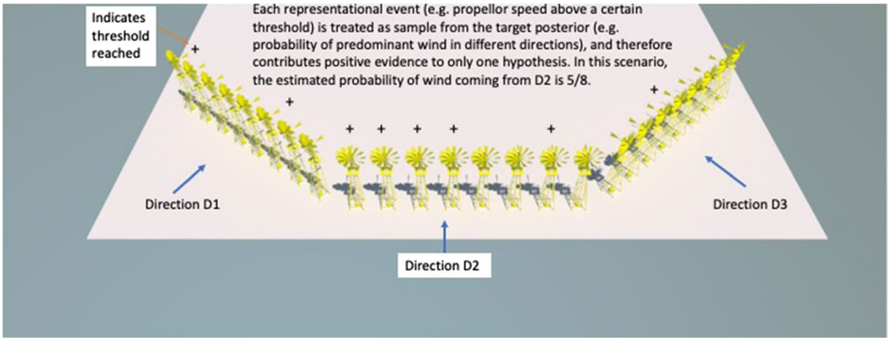

Fig. 3 Windmills used as a sampling system

(see Figs. 1, 2 and 3). Suppose you are interested in measuring the direction that the wind is coming from, using windmills that respond in a fairly reliable but also somewhat indeterministic way to wind coming in the direction they are facing; the windmills are also driven, albeit less strongly, by wind coming at non-perpendicular angles.

A representational event in this case is a windmill spinning at a certain rate at a certain time, and our goal is to infer information about the prevailing direction of wind from these events. Different wind detection strategies will differ partly based on the following choices:

Number of Representational Events: Does the strategy involve one representational event or many (e.g. many windmills, or many readings from a single windmill)?

Relevance of each Representational Event: Is each representational event used as positive evidence for the probability of a single hypothesis (e.g. a single wind direction), or as positive evidence for the probability of many hypotheses (e.g. many different wind directions)?

The models we consider are each the predominant way in the literature of developing a detection strategy based on different choices of number and relevance of representational events. Signal detection theory involves one representational event and one hypothesis (e.g. is the wind coming from direction D or not?). A variation on signal detection theory involves a single representational event, but we consider its relevance to many different wind directions. With probabilistic population codes, we have many representational events (e.g. the rates of many windmills pointing in different directions), and we consider each event as evidence for many hypotheses (e.g. the rate of one windmill contributes to my estimate of the probability of *every* individual wind-direction hypothesis). With sampling models, we still have many representational events (e.g. many windmill readings), but we simplify things by (among other things) considering each event as only positive evidence for *one* hypothesis (e.g. one wind direction). Let us now look at each such model in more detail.

1.1 Signal Detection Theory

To start with the simple case of the kind considered by signal detection theory, suppose you have only one windmill, and you are interested in using it to determine whether the prevailing wind direction is from the west (which we will define henceforth as "within 5 degrees of westward"). You might then face it in a westerly direction. You take a single reading of how fast the windmill is turning. You are aware that the windmill might turn in the absence of westerly wind (we are dealing with a noisy process), but the more strongly it turns, the more likely it is that the wind is indeed westerly. We might also assume here that you know that the average wind speed is fairly constant, so we can ignore this other source of variation in propellor speed.

In this situation, we have only two relevant parameters—which direction the windmill is facing, and how fast it is currently turning. We suppose that this enables us to read off a single probability for the hypothesis of interest; for example, we can imagine using a chart illustrating a monotonic function from propeller speeds to probabilities of westerly wind.

More specifically, signal detection theory theorizes this "reading of probability" in terms of two *likelihood* distributions (i.e. functions telling you the probabilities of signals given stimulus values), one for the probability of different speeds of spinning when wind from the west is *present* (target present) and one for when it is *absent* (target absent) (see Fig. 4). This gives us the *likelihood ratio* for target present vs

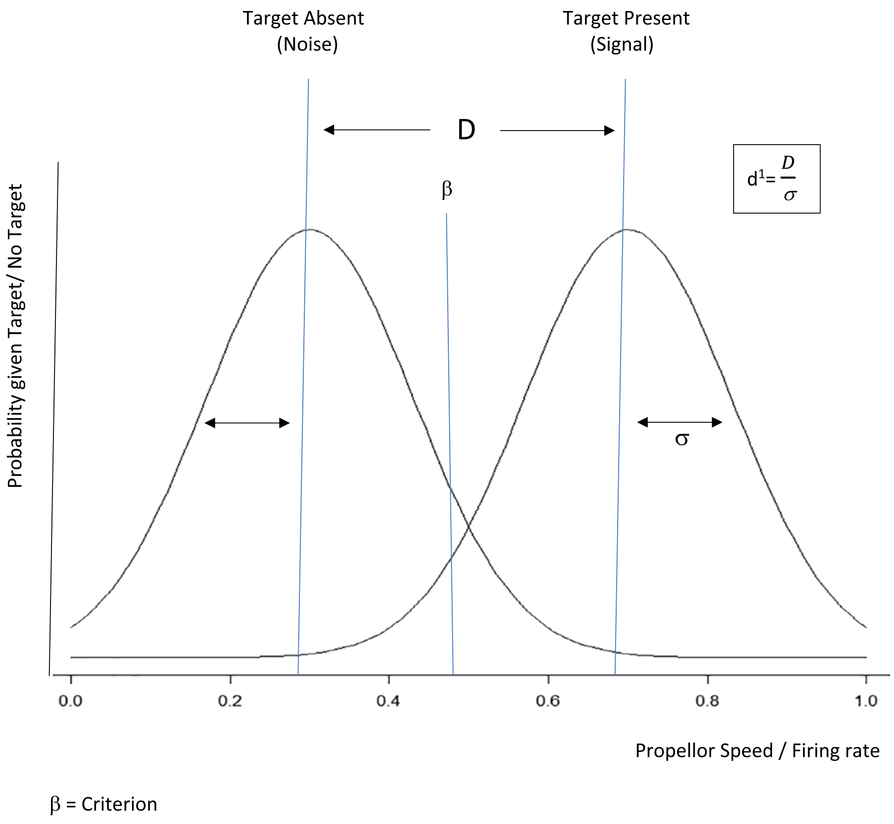

Fig. 4 Signal detection theory

target absent given a particular signal strength. If we assume a flat prior, this is proportional to the posterior probability – that is, to the probability of westward wind given the spinning of the windmill. If we include priors, then by Bayes theorem, we must also weight our likelihoods by the priors to calculate the posterior.

This modeling of how the signal responds to the target is referred to as the *sensory stage* of signal detection theory (Green and Swets 1966; Gescheider 1976). There is a further *decision stage* where we imagine a subject observing the windmill and making a decision as to whether wind is coming from the west, based on how strong the signal is. SDT assumes they use a threshold value, called a criterion. The problem of setting a criterion in a rational way can be given a Bayesian analysis, where we consider an ideal observer who is equipped with knowledge of the relevant likelihood functions and a prior, and a cost function that deals with the problem of how to trade-off between false positives and false negatives. For example, if it's very important to not miss westerly wind when it is present, then we might set a low, liberal propellor speed as our criterion, even though there is a fairly high chance that this signal could also occur when wind is coming from a different direction.

One of the key insights of SDT is that we can measure how good the signal is for detecting the target in a way that is independent of this decision stage. This is given by the signal-to-noise ratio, that is by how far apart the target-present and target-absent distributions are (usually given by a measure "d-primed" that is proportional to the variances of the distributions and the difference between their means). It's true that if we have a very conservative criterion for westerly wind, then we may miss many more cases of westerly wind than an individual who sets a very liberal criterion. But that doesn't mean we are worse at detecting westerly wind in the relevant sense.

We have been describing the result of setting a criterion – the "decision" – as a binary state, where the outcome is "west wind" or "not west wind". But the result of the decision could also be a state that reports on a 1 to 10 scale how probable it is that the wind is coming from the west, and/or how probable it is that it is *not* coming from the west, or choosing an action that is appropriate given how probable the target state is. Thus the decision could reflect the posteriors in a more systematic way than we get from a mere binary decision, a point we will return to later.

In the neuronal case, neurons and their activation strengths are analogous to the windmill and the speed at which it turns. Instead of thinking of a neuron as tuned to the presence of wind, we can think of it as, say, tuned to the presence of a predator. In this case, the signal is a neural firing rate (or a functionally analogous neurally implemented signal) rather than a rate of spinning. The continuously variable strength of the signal functions to indicate the probability that a predator is present. An appropriate "decision" to behave in a certain way can be based on the signal strength, for example the agent might act with caution with a moderate predator signal, or run away when the signal is strong.

1.2 Probabilistic Population Codes

Now suppose we want more detailed information about the probabilities of different wind directions, not just westerly wind. Even with just one west-facing windmill, we potentially have information about other wind directions—for example, since southwest wind also drives the west-facing windmill, if the windmill is spinning a certain amount when we measure, this might increase to some degree our confidence in south-westerly wind. Indeed, if the westerly windmill turns at low speed, then this might even be a *better* indicator of south-westerly wind than westerly wind. It might even be a better indication of southwest wind than the reading from a southwest facing windmill (Pouget et al. 2000 note that firing rates of neurons whose tuning curves are slightly offset from a target stimulus value can be better evidence for that value than neurons directly tuned to it, because they have a narrower band of responses that favor that value).

This "SDT with many hypotheses" extension becomes interesting with the further development of having *many* windmills which we can set up facing in different directions – a "population" of windmills (Fig. 2). Here's how we would use them as a "Probabilistic Population Code" (Ma et al. 2006; Ma 2010; Ma and Jazayeri 2014). Now we are not just interested in whether wind is coming from the west, but also interested in the probability, for *every possible* wind direction, that it is the prevailing wind direction. We use a single reading of propellor speeds from our bank to determine this. We consider each wind direction in turn, and use *all* the propellor speeds as data for that direction. For example, if I am interested in westerly wind, then I get good evidence not just from the reading on the westerly windmill, but also from all the windmills that are facing in directions not too far from the west. In principle, if I know, for every windmill, the probability of different windmill speeds given westerly wind vs other wind directions (the "likelihoods"), I can calculate the posterior probability of westerly wind. In probabilistic population code models, the response of each windmill when wind from each of the possible directions is present is plotted in a "tuning curve" whose peak corresponds to the preferred wind direction of a particular windmill. PPC models tend to separate a sensory stage and a decision stage where, like in SDT, we imagine an ideal Bayesian observer who combines the windmill readouts with knowledge of the tuning curves (i.e. likelihood distributions) (and possibly a prior) to calculate the probability of wind from a particular direction. This calculation could be completely intractable, unless we make some assumptions about the form the likelihood functions take. A central theoretical result of the literature on population codes is that if we assume (realistically) that noise comes from the family of "Poisson-like" distributions (which have the feature that their mean and variance are proportional), then we can calculate the log likelihood of a given wind direction as a fixed *linear combination* of the different windmill speeds. From the log likelihood we can, in turn, calculate the likelihood with a fairly easy mathematical operation. This has some big computational advantages we will discuss below. In some PPC models (Jazayeri and Movshon 2006), this leads to the postulation of a *second stage* population code, where each neuron's firing rate linearly represents the log likelihood of a single hypothesis (by analogy, picture each windmill only representing the likelihood of a *single* wind direction).

1.3 Sampling Models

Even with the Poisson noise assumption, the computations involved in standard models of PPC are quite intensive, because every neuron in the code is data for every hypothesis. Sampling models propose a simpler one-hypothesis-per-representational-event picture of how neurons represent probability. There are a variety of different models of this kind in the literature. We will discuss three ways of developing a sampling model, a *synchronic* version, and two *diachronic* versions.

The basic idea behind sampling is to simplify the problem of computing over probability distributions by replacing them with a set of representative samples, which can be defined as items that are used to *stand in* for the different hypotheses in the state space (in our example, hypotheses about what direction the wind is coming from). A theory will postulate a *generative process* that produces samples with probabilities that (ideally) will match the posterior distribution we are interested in. So, for example, if we want to know the proportion of white and black balls in an urn (and so the probability of randomly pulling out a white ball), the generative process of pulling out a single ball and then replacing it will produce samples with representative statistics; the more times we sample, the higher chance we have of a representative set of samples. As we will illustrate momentarily, further possible features of theories are that (1) *weights* on samples (e.g. the firing rate of a neuron that stands for a particular state), are used in conjunction with *numbers* of samples to play the representative role—rather than summing samples directly, we sum them based on their weights; (2) we deal with large or infinite state spaces by *pre-selecting* only a small number of states to get potential samples from; (3) we postulate an *algorithm* as part of the generative process, whose iterations generate a gradually more representative set of samples (e.g. a particle filter (see below)).

In one kind of sampling model, samples are produced *synchronically*—e.g. a synchronic firing pattern across a set of neurons might be treated as a set of samples (Lee and Mumford 2003, Hoyer and Hyvärinen 2003). By analogy, consider again the bank of windmills used to measure different wind directions. Whereas in PPC each windmill was used as data for every direction, we now only treat each windmill reading as giving us information, in the form of a potential sample, about *one* wind direction. One way to do this is to set things up so that each windmill is dedicated to producing samples for only one hypothesis (e.g. the direction the windmill is facing). On that version (Fig. 3), since we have only finitely many fixed-direction windmills, we can only get samples from finitely many of the infinitely many wind directions, those that we *preselect* (confusingly, one could also naturally describe this pre-selection process as "sampling", but here merely setting things up to get information about, say, northerly, north–north-westerly, and north–north-easterly wind is not the same as actually getting this information, which is what we mean by "sampling"). We may also choose to make things more simple by preselecting only a subset of all the windmills that we have (e.g. 160 of 1000). These pre-selected directions and windmills might be initially chosen at random, or strategically according to our prior sense of where the wind might be. Once the windmills are set, the rest of the generative process is simply the wind producing a set of windmill speeds, which then can be interpreted as representative samples in one of two ways. To get a "number of samples" style of representation, picture that we put 160 windmills between each of 16 compass directions, so 10 like-facing windmills in each direction we want to sample from; and then take the generated number as those whose speed goes above a certain threshold—so e.g. we might get 8 samples in a northerly direction, and 2 samples each in a north–north-westerly and north–north-easterly direction.

If weights are used as well (see Lee and Mumford (2003)), then instead of using a threshold, for each direction we sum across our 10 windmills based on their propellor speed, to get a total *sample strength* in that direction. If the generative process is set up nicely, the proportions of these numbers/strengths are approximately proportional to the relative probabilities/likelihoods of the relevant hypotheses (so e.g. the probability of north wind is 8/12 in our example). Our samples therefore give us a fairly literal depiction of the probability distribution we are interested in (e.g. probability of wind directions given readings), albeit a discrete approximation of it. The more samples we have, the better the approximation (and if we are using weights, how *well-calibrated* our propellors are also matters). In cases where we pre-selected, we must accommodate the fact that we ignored many wind-directions that we also want to assign probabilities/likelihoods to; this can be dealt with by a rational interpreter of the samples through *interpolation*—e.g. assuming that the density function varies smoothly between the hypotheses we have sampled on (or similarly, one can think in terms of *sample density* in a region of hypothesis space as a proxy for probability density in that region).

Other models are diachronic, involving samples produced over a period of time. In one kind of diachronic model, we have an iterative algorithm that produces a better and better set of synchronic samples (so it's an extension of the generative process). For example, suppose at the first step, we pick oriented windmills fairly randomly. By reading their speeds, we discover that some are probably facing away from the wind, and so not telling us much that is interesting about wind direction, while others are better predictors. So we *resample*—we redistribute our pre-selected windmills in a biased way, such that they are more likely to be positioned in directions where we had higher wind-speed readings in the previous round. We now take another set of readings and repeat the process. As we continue, our samples get better and better tuned towards the direction that the wind is coming from.

This is a simple version of an algorithm known as the *particle filter*. It has been used to model how hierarchical bayesian inference could be achieved in the visual system (Lee and Mumford 2003). The system is modeled as a series of levels, each representing features at a different level of abstraction. At each level, we have a set of weighted samples (known as "particles") representing a distribution over the stimulus feature represented at that level. At each iterative step we "resample" based on both the likelihood of these samples given bottom-up evidence from the level below, and a prior generated from the level above.

There are also diachronic models where our representative samples do not exist simultaneously at all, but rather are produced in a temporal sequence (Fiser et al. 2010, Orbán et al. 2016). Suppose we have just one windmill but in this example, the windmill is capable of rotating. It is pushed in the direction of the prevailing wind, albeit in a very noisy indeterministic way, due to the non-uniformity of the wind itself, and the rusty bearings inside it. Rather than measuring its speed, we simply measure what direction it is facing at a given time. If I make, say, 100 measurements over a period of time, I would expect a larger proportion of measurements around the prevailing wind direction than other directions. As before, if things are set up very nicely, there could even be a simple linear relationship between the relative probabilities of wind directions, and the relative numbers of readings (or "samples") in those directions. So if out of 100 measurements, I find that the windmill is oriented west 60 times, north-west 20 times, north 10 times and south 10 times, and things are set up nicely, I am in a position to know that the probabilities of wind directions given these readings are directly proportional.[4]

[4] One can also have a similar model involving a population of neurons at a given time representing a single stimulus configuration, and the population cycling through different representations over time at a rate proportional to the probability of the stimulus configuration (Orbán et al. 2016)). Relatedly, one can have a synchronic analog of this "cycling neuron" (or "rotating windmill") model, where we have many copies of the neuron/windmill simultaneously, each capable of indicating the full range of wind directions but only indicating one at given time, with probability proportional to the probability of that wind direction. Again, number of samples (or sample density), can then be taken as a proxy for the probability of wind from a given direction (or range of directions).

Sampling models rely less on a decision stage than other models, in that they provide the most literal rendition of the probabilities of states, which therefore does not require any further "decoding": the system can use individual probabilities and do bayesian calculations on them by directly using the sample numbers/weights. They also do not require Poisson-like noise, as with PPC, but they do make the strong assumption that the distribution of samples will tend to match the probabilities of states. In this sense, they "make a virtue of noise" (Hoyer and Hyvärinen 2003).

In what follows we will not assess the adequacy of sampling or of the other models we have been describing. We take the issue of accuracy to be largely empirical. That said, one might already be worried from this presentation that the appeal to ideal Bayesian observers armed with likelihood functions they use to interpret signals in a rational way is already highly idealized or metaphorical, and so be unclear what the literal sense is in which these models postulate probabilistic representations. The *modelers* can certainly understand the function of the system in terms of the probabilities encoded by neuronal activation patterns, but what the modelers do is not what the system modeled does.

A common way to develop this worry is by appeal to Marr's famous distinction between levels of explanation (Marr 1982). Because Marr's distinction is widely known and cited we will not rehearse it here. The basic point is that skeptics of probabilistic representations in perception can argue that while the function of perceptual systems can be understood in a Bayesian way at the computational level, such systems do not genuinely use probabilistic representations at the algorithmic level. The function of many systems, including plants, can be understood in Bayesian terms without using a Bayesian algorithm (Block 2018). In this context, models of sensory and perceptual processing that posit probabilistic representations have been criticized for failing to use a control class – an alternative algorithm that can potentially account for just the same kind of Bayes optimal performance without using probabilistic information (Bowers and Davis 2012).

Although the probability distributions in SDT, PPC and sampling models can play a role merely as part of computational-level *just-so* stories, it's also true that these models postulate states that could be probabilistic representations in a robust sense (the signals, populations of signals, and sets of samples), and postulate operations on these states that at least *approximate* the operations of the ideal agent. That's why they are interesting for our purposes. We think that if one is interested in what makes these states count as probabilistic representations, how they are actually used by the system, and in particular how they are used, *at least approximately,* in a probabilistically appropriate way in computations, is of central concern.

Bearing this in mind, we now turn to examining in more detail the crucial issue of the individuation of probabilistic states. It's useful here to distinguish two (interrelated) issues. First, a probability distribution is a structured mathematical object, and it is being used here to model or measure an internal state in a way analogous to how real numbers can be used to model physical quantities. This issue raises important measurement theoretic questions: how uniquely does the structure and role (perhaps including connections to the environment) of the probabilistic state determine a given probability distribution? Are there other kinds of mathematical objects (perhaps with less structure) that could be used equally well in this context? The second issue is this: there is more to saying that we are representing probability than merely that a certain kind of measure on hypotheses can be associated with a state. What makes that measure a *probability* measure, rather than some other kind of measure? Presumably, this has to do with it being used in a probabilistically appropriate way. But what exactly does this mean in the context of sub-personal probabilistic states?

Our focus here will be more on the latter issue, even though the former is also very important (for example, we would need to consider it to understand how much indeterminacy there is in the content of probabilistic states). We begin by approaching the issue we are interested in by comparing the individuation of probabilistic states with the individuation of credences.

2 Comparing Subpersonal Probabilistic states and Credences

Credences (aka degrees of belief), as they have been theorized in work in formal epistemology, can be understood as mental states which are, in a sense, probabilistic representations. Much work has already been done to address, for credences, the two issues just mentioned, so a comparison with our target states is a natural place to start.

Philosophical work on the notion of credence starts with the intuitive idea of believing something with a certain level of confidence, and then proposes a notion of credence that extends and refines the intuitive notion in various ways (thereby making it controversial whether we even have credences) (Pettigrew and Weisberg 2019; Sturgeon 2020). For current purposes, we will understand credences as having the following features:

(1) Credences are real functional properties of doxastic states.
(2) They are measurable on a ratio scale and have a maximum value (probability 1 or 100% confidence).
(3) They assign values to members of a state space.
(4) They interact with the subject's utilities to produce behavior.
(5) They are subject to rational norms (laws of probability, laws of updating, norms of decision theory, etc.).
(6) They are subject to functional constraints that correspond to these norms (e.g. the system must approximately obey the norms to count as having credences).
(7) They are a feature of attitude rather than content.

Let's unpack these ideas, starting with (1)-(3). Credences are a kind of belief, or belief-like state, which differ from all-or nothing beliefs by including an additional vehicle component that assigns a weight to the proposition represented by the belief (the representation of the content being the other component—e.g. a sentence in the language of thought). Models that postulate credences also require there to be a state-space—an (ideally exhaustive) range of alternative hypotheses to which different degrees of confidence are assigned. If we have a finite number of hypotheses, then we can assign non-zero credence to every hypothesis. But we can also deal with the continuous analog of a finite space, by treating confidences as holding primarily over hypotheses about *ranges* of values rather than individual values.

Classic work on credence in epistemology and decision theory, under the influence of seminal discussions by Ramsey and De Finetti, proposes that we ignore the phenomenal dimension of confidence (e.g. that feeling of being really certain about something), treating confidence in a purely functional, or even behaviorist way. When we consider *experiential* confidence later, we will reinstate this phenomenal criterion, although the phenomenal features we are interested in need not feel "confidence-like".

This classic work addresses the measurement theoretic issue, arguing that we can measure confidence on a ratio scale, not just an ordinal scale, so that we can make sense of claims like "I am twice as confident that the wind is coming from the west than the east", and that there is a maximum level of confidence (that we can think of as confidence 1 or 100%), so that, given the ratio scale assumption, individual confidences correspond to precise numbers between 0 and 1, or precise percentages (e.g. I am 67% confident that the wind is coming from the west). The models we are looking at also assume ratio-scaled probabilities—for example, ratios between numbers of samples represent probability ratios in sampling models. Exact credence-ratios for beliefs have been challenged as psychologically unrealistic; by contrast, the accounts in perceptual neuroscience we are considering explicitly tell us what grounds the analogous ratios (e.g. sample-ratios or firing-rate ratios), and so an analogous commitment is less obviously problematic.

Importantly, none of the candidate probabilistic-representations we will consider are beliefs or belief-like; in particular they do not postulate separate vehicle components for representing the state of the world and the level of credence. For example, the signal on which a decision is made in SDT is an unstructured state (akin to the speed of a propellor) which does not involve any conceptual structure or other complex representational structure, such as imagistic structure. With the more complex states we get in PPC and sampling models, we have more structure and therefore potentially more *representational* structure, although the structure mostly functions to make explicit the credence-like properties of the state, and *not* to provide a structured representation of the propositional content involved (e.g. we do not have separate elements that mean "wind" and "west" combined into a sentence). In this sense, these perceptual states are rather similar to single-neuron *detectors*. One could in principle augment these accounts with a more explicit structured representation of the content linked to the credence like elements, but we are not aware of any models that do this.

A further property of credences is that they are subject to rational constraints or norms. The least controversial norms correspond to the laws of probability. For example, one law of probability is that the probability of the conjunction of two independent events is the product of their probabilities. Correspondingly, the rational confidence to have in the conjunction of two independent events is the product of confidences in the conjuncts. Many also believe in further rational constraints, for example constraints for updating credences based on new evidence, such as versions of conditionalization, or constraints on prior credences (e.g. the principle of indifference). In decision theory, there are also proposals for rational constraints on how credences ought to interact with utilities (the analogue for credence of how much you like something), to produce rational behavior. Although it is extremely unrealistic to suppose that agents always conform to these rational norms, they can nonetheless be part of an account of what individuates credences, in the sense that they generate functional constraints. For example, one might suppose that a system has to at least *approximately* satisfy these norms to count as having credences (a credence-involving version of the intentional stance has to be useful and explanatory), and it can only do this if it functions in certain ways.

The extent to which similar norms also individuate subpersonal probabilistic representations in a similar way is a central issue here. Plausibly, and as we will argue below, to count as a probabilistic representation in a robust sense, a state ought to be used inferentially in a way that respects similar probabilistic norms. For example, below we will discuss the ways in which PPCs can be used in probabilistically rational inferences to produce veridical percepts. However, there may be important limits to the analogy with the role of credences. A key point is that an appeal to the way utilities and credences combine to produce rational behaviors probably won't always easily transfer to perceptual probabilistic representations. For example, a traditional way of understanding subjective probability is in terms of *propensities to bet* (De Finetti 1937/1964). The relative confidence that one has in two propositions should manifest itself in what one takes to be fair betting odds for these propositions—for example, if I am 0.6 confident that the wind is coming from the west, then I should be prepared to accept 60/40 odds or better from a bookie on this (assuming that I value money in a certain way, that I don't disvalue taking risks etc.). Since our visual systems don't accept odds from bookies, and do not do anything analogous (perhaps a less obvious assumption), then such an account will not apply. Instead, the analog of rational behavior will be the output of a perceptual task, as it might be modeled at Marr's computational level. Weights found in perceptual systems can be interpreted as probabilities provided they take part in computations that are then seen to make rational sense, given the computational task the system is trying to perform. If the system is ultimately aiming at an accurate representation, such an account will be analogous to the account of credences we would give for a "pure enquirer" who forms beliefs based on probabilistic reasoning, but who does not act in the world and so has no use for a cost function.

A cost function *can* also play a role in modeling perceptual processing, but typically in a way different from the "betting propensity" kind of model. For example, we saw earlier that a cost function can be implicit in the way that the system trades off between false positives and false negatives. It should also be noted that given the intractability of many bayesian computational problems, rather than directly following the optimal rational norms, the system may be following heuristics that merely give results that approximate such ideally rational rule following, perhaps only in a limited range of contexts. We allow that following such heuristics can still be a way to "use weights as probabilities"; since some of these heuristics may have no analogs at the personal level, this is another potential disanalogy with credences.

Traditional approaches to credence like De Finetti's are notoriously behaviorist; for both credences and sub personal probabilistic-representations, we may prefer a *functionalist* view. For credences, an example functionalist position might say that credences and utilities are real aspects of internal states that are functionally individuated in terms of their interacting via *expected utility calculations* to produce choices between behaviors. In this way, credences are not just summaries of behavior or behavioral dispositions, but real internal causes of behavior. Similarly, the probabilistic-states that we are considering are real causally active internal states, that may be individuated by computations comparable to, but perhaps different from expected utility calculations.

We will remark in passing here that Icard's recent account of subpersonal subjective probabilities as sampling propensities doesn't meet this functionalist desideratum (Icard 2016). Roughly, Icard treats a generative process in the brain that produces samples as implicitly having certain subjective conditional probabilities, provided it behaves *as if* it represents those conditional probabilities in the way it produces samples (i.e. the long run sample statistics match the represented probabilities). Treating implicit knowledge in such a behaviorist way has long been a temptation, but there are also theoretical reasons for pushing for a more functionalist account. As Davies (1995) has argued, ascriptions of implicit knowledge can be understood as putting constraints on the causal structure of the system, and therefore offering a deeper explanation of *why* the system has the behavioral dispositions it has (e.g. its behavior isn't just generated by a giant lookup table). Another way to put the same point is that unless our theory generates such constraints on the causal structure of processing, it is really only a computational-level account, at best explaining the input–output function in a teleological way by comparison with an ideal standard. Future work on the implicit representation of conditional probabilities could try to articulate such constraints in terms of features like neural connection strengths, to obtain a more realist-functionalist algorithmic-level perspective. Here we set aside implicit conditional probabilities to focus on the explicit (i.e. occurrent) probabilistic representations exemplified in our models.

Credence is also traditionally viewed as an aspect of the attitude of a belief rather than the content (Titlebaum 2019), although there are exceptions (Moss 2018). There is a difference between having an 80% credence in a westerly wind and believing that the probability of a westerly wind is 80%. One reason for drawing this distinction is that beliefs about probability require the concept of probability, but one could have confidence levels without this sophistication (Gross 2020, p. 383). Does this distinction apply to perceptual states also? We don't think it will do much theoretical work in the domain we are interested in, because we are not dealing with structured propositional states of the kind that could explicitly involve the concept of probability. That said, it is more natural to regard our representational weights as aspects of attitude rather than content, just because the content (i.e. the state space) is determined externally by what our neuronal signals respond to in the environment, whereas our subjective probabilities are an aspect of the *internal functional role* of the states that can be treated quite separately from whatever gives them their content. We will further develop our understanding of this probabilistic role in the next two sections.

Taking stock to conclude this section: perceptual probabilistic-states may differ from credences in:

(1) Not being belief-like, in particular by not explicitly representing their contents as a separate structured component.
(2) Not interacting with utilities to produce rational behavior according to the norms of decision theory.
(3) Not being functionally individuated by exactly the same set of rational norms as credences.
(4) Not fully sustaining the same distinction between probabilistic contents and attitudes.

That said, our probabilistic states still share two basic features with credences: they involve parameters (e.g. firing rates) measurable on a ratio scale that play the role of representing a probability distribution over a hypothesis space, and they play a role in explaining the functioning of visual processing that is analogous to the role that credences play in explaining the inferences drawn by a rational agent trying to infer beliefs about the world from evidence. There may be some differences in the kind of inferential processes that occur subpersonally vs in an ideal rational agent (e.g. the use of fallible heuristics), but the basic principle of states being appropriately interpreted as representing probabilities due to their appropriate use in an inferential process applies in both cases.

3 Thin Probabilistic Representations in Perception

We now consider our three empirical models from a different angle, through the philosophical theory of mental representation, considering how a variety of accounts of representation can be adapted for our purposes. We will first focus on thinner notions of representation, and then discuss a more robust sense in which SDT, PPC and sampling may introduce probabilistic representations in perception.

Because they can be successfully used to establish probabilities, the windmills and neurons as modeled in SDT, PPC and sampling might be said to "carry probabilistic information". One way to develop this idea is in terms of the probabilistic *informational content* theorized by Skyrms (2010), inspired by Shannon's notion of information (despite the fact that this is *not* typically thought to be useful in understanding the content of a signal). Skyrms' notion differs from the familiar notion of propositional content, in that propositional content asserts that a particular state of affairs obtains (ruling out others that do not obtain), whereas informational content communicates how probable different states of affairs are. Specifically, the informational content in a signal consists in how the signal *affects probabilities* (2010, p. 34), and can be described by a vector whose entries correspond to states in the state space. If a given signal "moves" the probability of a state from 1/10 to 9/10, for example, then the signal carries probabilistic information about the state, which can be represented as an entry of 0.8 in the vector slot for that state. (One could also envisage variants of Skyrms' idea where the entries in the vector are instead posterior probabilities (Godfrey-Smith 2012), or likelihoods).

A spinning, westward-facing windmill carries informational content about the probability of westward wind because the spinning corresponds to an increase in the probability of westward wind being present. *Ditto* for neurons that are tuned to certain environmental elements and increase their firings in the presence of the elements. *Ditto* for groups of windmills, or groups of neurons. Skyrms' notion of informational content, however, does not require that this probabilistic content be in any sense *used appropriately* or *made explicit* to the system itself, and this is typically seen as an important constraint on a thicker notion of mental representation. Informational content is ubiquitous in nature. Neural activity "moves" the probability of states of affairs that the neurons are presumed to be tuned to, but it also moves the probability of the presence of serotonin and melatonin in the brain. Yet we don't generally regard the firing of neurons as signals of serotonin levels. A state that merely affects probabilities is certainly a useful causal intermediary, but, in Ramsey's terminology, it is not thereby also a state that is playing a *representational role*.

Similar considerations come up when we add the teleo-functional idea that in addition to having informational content a signal should have *the selected function* of carrying probabilistic information in a system (Dretske 1997). Although a neuron's firing response and a windmill's speed of spinning may be designed, and selected for co-varying with the increased probability of a certain state of affairs, it is questionable whether this delivers a robust notion of representation. States that have evolved to co-vary with environmental parameters are ubiquitous in nature and appear in plants, and in bacteria. For example, a cell might respond in a variable way to the concentration of a chemical cue in a way that a theorist could fruitfully model in terms of a probability distribution over possible sources of the cue; the chemical receptors in the cell thereby may have the selected function of carrying probabilistic information, but they are not probabilistic representations in any robust sense.

For these reasons, we are motivated to look for further conditions. On the question of what, in general, makes a state a representation with a certain content, Ramsey (2007) makes the useful move of suggesting a divide-and-conquer strategy, whereby we distinguish issues of representational-status—what makes something deserve the name "representation" in the first place—and the issue of representational content—given that it is a representation, what makes it a representation of a green plant rather than a piercing trumpet? We propose to follow the same pattern for probabilistic representations. Our discussion is more focused on the questions of status, although we do not suppose that the issue of status and the issue of content are independent problems.

We think that the motivations for accounts of representational status transfer over well to the probabilistic case. A representation doesn't simply carry information, but rather functions to carry information in a systematic way and in an exploitable form—an idea that can be developed in different ways (see Shea 2018 for extended discussion). The information is there "for the system" in some sense, and not merely for an external theorist trying to understand the system. Similarly, to get a richer notion of probabilistic representation, we want a sense in which probabilistic information is not just relevant to the theorist analyzing the system, but is there "for the system" itself, as it were (Shea 2014 concurs, but adopts a rather different approach). We develop this idea in the next section.

4 Robust Probabilistic Representations in Perception

One indication that probabilistic content in the above sense is not "there for the system" is that it is entirely a matter of the objective statistical relationship between the state and the world, and need not in any way be reflected in the internal structure or functional role of the state. By contrast, subjective probability surely has much more to do with the internal use of a state.

One familiar response to this observation is to introduce a consumer, or user—possibly a separate component of the system—that exploits the information in an appropriate way (Millikan 1989; Shea 2018). The obvious danger with this strategy is that the user is treated as a sort of homunculus. For example, in SDT and PPC we imagine a Baysiean observer who, in the decision stage, performs Bayesian calculations to extract the probability of certain elements in the environment from the firing patterns of neurons. But this is certainly an idealization. In the absence of a user of this sort, what kind of use would we need to see *in the system* to judge that it is indeed exploiting probabilistic information?

One option is to take a look at the actions that the signal prompts. So for example, it might be that fleeing is appropriate (e.g. adaptive) when the chance of a predator is above 50%, and that fleeing is prompted by a signal above a certain threshold. The trouble here is that this "use" of the signal is so thin that it is unclear we are in different territory from examples like cells responding to chemical cues. It's a mere behavioral switch.

As we see it, there are several related ways to get a more robust kind of use of a probabilistic state. One is if the signal is used in a wider range of probabilistically appropriate computations. Another is if the content is used in a more flexible context-dependent way by the system—particularly, in a way that takes into account what *else* is represented about the world and the organism's needs. Another is if the consumer includes a subject or system that uncontroversially has credences or credence-like representations, and they form these in an appropriate way in response to the probabilistic -states (so we really do have something a bit like a homunculus consuming the representation).

To illustrate the first two constraints, and to stick with a SDT-style example, suppose an organism has a detector signal for a food source like a fruit. A crude use of the signal would be a seeking behavior if the signal exceeds a certain threshold. A more sophisticated use might enable the organism to take into account the prior probability of a fruit present in this context, how important the food source is given their current needs, and allow graded variety in behavior depending on the posterior chance of fruit—e.g. a low signal might merely prompt an attempt at getting a better look, a high signal might prompt a fruit-fetching behavior—all provided a fruit fits the organism's current needs. An even richer representational use might also allow the system to compare and integrate this information source with other information sources regarding the fruit-status of the target (e.g. smell), and to reason in more sophisticated ways about what behavioral options are available—taking into account, for example, the different possible uses of the fruit—e.g. feeding it to offspring—or planning to retrieve it later when it is more needed.

What these considerations already suggest is that there is not a clear dividing line between crude relay switch type uses of probabilistic information, and more sophisticated representational uses—it is a matter of degree, depending on the extent to which a state is used in these richer ways. Our sub-personal perceptual probabilistic- states are likely to be at least somewhat further towards the relay-switch end of the spectrum than personal-level credences, because in general perceptual states in sub-personal systems are used in a less flexible, more stimulus dependent way, and may not be able to take into account utilities or as rich a range of sources of relevant information.

As for the third way of getting a more robust kind of use of a probabilistic state, if we are dealing with a human-like subject that has credences, then a perceptual signal potentially gets a kind of derivative status qua probabilistic representation from its role in generating credences. So for example, if our signal is a graded fruit perception that leads to different *degrees of belief* that a fruit is present, then it is being used by a consumer in a way that is akin to a person reading a windmill – that is, it is being used as a representation in a fairly literal, quasi-homuncular way.

Even with these more robust constraints in place, there is a danger of incorrect classification, a point on which we will return when we discuss experiential credence. There are examples where we have a variable weighted representation, but its primary role is simply to indicate the magnitude of an external feature (for example, distance, intensity or brightness), and not to represent probability. If the magnitude of the feature *correlates* with the probability of a relevant hypothesis (e.g. the more intense a cue is, the more likely it is that the source of the cue is dangerous) then the represented feature could be used contingently like a probabilistic representation. For example, the strength of an odor from a food source might tell an organism how likely it is that it is not safe to eat. If the organism acted in such a way as to be sensitive to the probability of poisonous food, then we might interpret this case as the organism being guided by a probabilistic state about the chances of poisonous food. But such a state might simply be formed *on the basis* of a perceptual state that is not itself probabilistic. This would count as a probabilistic *use* of a perceptual state, but there would only be a weak sense in which the perceptual state itself is a probabilistic representation.

Examples of this kind suggest that to have a clear case of probabilistic representation in perception we would want to see a perceptual state that systematically *and exclusively* carries probabilistic information, where the information itself is being appropriately exploited in the system. In the example we just described, the representation of the intensity of a smell does *systematically* co-vary with the probability of the food being dangerous, but it does not *exclusively* play a probabilistic role (intensity does not always correlate with poisonousness). Exclusivity will be important below in considering conscious perceptual probability.

As for *systematicity*, it is a central feature of *analog representation* (Beck 2019). Whether we are dealing with a feature space (e.g. spatial distance) or a space of probability density functions, if differences in the target space are marked in a systematic way in our representational scheme, this makes the scheme more apt for performing computations that appropriately reflect the structure of the space. Arguably, the kind of systematicity that we see in our models of probabilistic representations is *structural resemblance* between the target and the representational vehicle, of a kind that is aptly thought of as a kind of *analog representation*. Moreover, the computational uses we have been describing are implemented as a form of *analog computation*. Let us elaborate.

Analog representations, as we understand them, are representations that use a variable quantity in the vehicle (call this the "representational space") to represent a variable feature parameter (e.g. spatial distance, duration) (the "feature space"), where these representational contents are given by a structure preserving mapping from the feature space onto the representational space. In many cases, this is not a simple linear mapping – e.g. in classical psychophysics, features like brightness and perceived weight are thought to involve log or power psychometric function. But nonetheless, the mapping preserves structure in such a way that it can be used in an informative way to perform truth-preserving computations, or to guide action or decision making in an appropriate way. For example, if I have a monotonic psychometric function for a 1-dimensional quantity like spatial distance, I can compute which of two spatial distances is larger by computing which of their representations has a larger value. In what follows, we will take (1) structural preserving mapping, and (2) appropriate use/exploitation of this mapping in computation / downstream process, as the two hallmarks of analog representation.[5] These are very much not independent constraints; as many authors have pointed out, the existence of many structure-preserving mappings from vehicle to target is often completely trivial. What matters is that a particular structure-preserving mapping is actually made use of by the system in the kind of appropriate ways discussed above.

The idea can be extended to probabilistic representation in a fairly obvious way. Instead of a space of features, we have a space of probability density functions, and we are interested in structure preserving mappings from this space onto representational spaces, such as spaces of vectors of neural firing rates, or just individual firing rates. Because the points in these spaces are themselves structured objects (unlike in the case of simple features), these mappings will also typically require that individual probabilities in the density function are mapped systematically onto relevant

[5] Some authors in the philosophical literature describe a notion of "structural representation" where the mapping that semantically individuates the representation is from concrete objects and events and their relations to relations between representations in the brain (e.g. Shea 2018). This is not the notion of analog or structural representation we are operating with here, where the mapping is from members of a space of stimulus *features* to representational features.

aspects of the representational object, such as the firing rates that form representational vectors (although this need not be in the simple form of a linear or monotonic function from probabilities to firing rates, as we will see below); so in a sense we have systematic mapping at two different levels here. Furthermore, we can appeal to the ways in which a mapping is appropriately exploited (e.g. in the ways described above), selecting it as a content for the analog representation. For example, there may be a range of computations on an analog representation that respect the relevant norms of probability given a particular mapping onto a probability density function.

In what sense do the probabilistic states in SDT, PPC and sampling models involve (or not involve) such analog probability representations? One thing to reiterate here first is that none of the models include any kind of analog representation of the *worldly states*, only of probabilities or other probabilistic parameters. So for example, we have an analog representation of a certain spatial distance having a certain probability only in the sense that we have a variable parameter encoding *probability*, but not one encoding spatial distance in an analog way. The models are typically silent as to how these other parameters, such as spatial distance are encoded. They could be coded in a non-analog way, e.g. by different neurons representing different distances. If a separate explicit representation of distance (e.g. an analog one) was added to the analog probability representation, then we would have a state much more akin to a paradigm credence – but none of the models postulate such a further component. The models tend to just assume we start out knowing what worldly elements neuronal populations are tuned to.[6]

Let's assume to begin with that the variable signals postulated in our models are variable physical quantities like firing rates, and so this aspect of the "analog representation" model is satisfied. The issue then is the sense in which the models involve an appropriately exploited structure-preserving mapping onto this physical quantity. This "signal as variable physical quantity" assumption will then be revisited below.

Bearing this in mind, let's look at the models. Start with SDT. Again, the way the theorist typically thinks of this is in terms of separate stimulus-to-signal likelihood functions for *target present* and *target absent* (see Fig. 4). For each signal value, we can compare these likelihoods, by taking their ratio. Target present is more likely than not when the ratio gets above 1 (i.e. above 50/50). If we assume a flat prior, then this ratio is linearly related to the ratio of posterior probabilities of target present / target absent. This means we can also take each signal value to correspond to a posterior probability of the target being present – the stronger the signal, the more probable it is that the target is present. Is the signal therefore an analog representation of a likelihood ratio, or of target posterior?

From what we just said, we clearly have the structure-preserving mapping onto a variable physical quantity, but is it exploited or used appropriately? Here we simply reiterate what we said earlier—basic binary decision making given a criterion is a

[6] One of us finds this feature of the models to raise questions about their aptness to adequately explain how distal conditions are represented given *underdetermined* stimulation. It seems that the models assume rather than explain what distal elements are represented by populations of neurons.We do not have space to expand on this point here.

very thin use, but we can imagine ways in which a signal could in theory be used in a more richly representational way, although this is not typically modeled very extensively in classic work on SDT.

This is less of an issue for probabilistic population codes, to which we now turn. Here recall that we have an initial stage with a bank of neurons with different tuning curves, responding with poisson-like noise. We can then calculate log likelihoods of different stimulus values as weighted sums of firing rates, potentially leading to a second stage where we have a population code where each neuron directly represents the log likelihood of a different hypothesis (Jazayeri and Movshon 2006).

If we have this second stage direct representation of log likelihood, this satisfies the structure preserving property of analog representation in a clear sense. If we assume a flat prior distribution, then this is also proportional to the log posterior distribution, so we have structure preserving mapping of posterior probability in a straightforward sense also. What about the use criterion? There are possible computational uses of the log-likelihood distribution that respect the norms of probability. We can calculate maximum likelihood by looking at the peak of the curve. We can compute the likelihood ratio of two hypotheses by subtracting their log likehoods (i.e. subtracting firing rates), and we can multiply probabilities by adding log likelihoods (i.e. adding firing rates). This means that computations like integrating distributions from different modalities, integrating information coming in over time along a single channel, or calculating the probability of functions of variables (e.g. the probability that two variables sum to a certain value), all can be done with fairly simple linear operations on firing rates – this is the big selling point of this scheme (Pouget et al. 2013; Ma and Jazayeri 2014). Furthermore, these simple operations on firing rates (addition and subtraction) are paradigms of analog computations; so an analog implementation is a major motivation for the theory.

What about the first stage population code? If this is to be thought of as an analog representation of likelihood (or log likelihood), it is so in an interesting indirect sense. It is not that each firing rate represents the likelihood of a given hypothesis. Instead, we can think of the likelihood function as a sum of weighted basis functions (a minimal set of functions such that every other function can be expressed as a sum of these functions), with each neuron representing one of these weights (Pouget et al. 2013). That is: relative to a set of basis functions, a given likelihood function is a *vector* representing the weights of the different functions in the sum; each neuron is associated with a different basis function, which can be thought of as representing, for each stimulus value, this neuron's importance in the calculation of the likelihood of that stimulus value at the decoding stage – i.e. roughly, how much that neuron responds to each stimulus value. The firing rate then gives the weight of that basis function in the calculation of the likelihood function.

Given this change of basis, operations on probability functions correspond to transformed operations in this alternate vector space. The fact that we can use such transformed operations is what makes this still a "structure-preserving mapping" in a sense still in the spirit of the idea of analog representation; indeed, this kind of point was already true in the simpler case of the second stage population code, where multiplication and division were transformed into addition and subtraction by taking the logs of our probability functions. Now, since there is a *linear* transformation from our first-stage weird basis functions into the second stage log probability basis functions, multiplication and division on probability functions at our first stage *also* correspond to addition and subtraction on these less familiar vectors. So all the computations we could do in a simple linear way *after* we got to the second stage, when we had log likelihood directly represented, we can *also* perform before we decode to log likelihood. For example, we can integrate probability distributions across sensory modalities, or integrate information over time using simple linear operations. So that means that most of the computational "uses" we could make of the log likelihood representations we can also make of these funny representations – we don't need to decode them first.[7]We have spelled this out here to illustrate how the "structure preserving" mapping involved can be far less obvious than, say, a simple linear map from probabilities onto firing rates.

Now let's turn to the sampling models. Recall that here we have a process that aims to produce a set of representative samples from a posterior distribution. They can be "representative" either in the sense that there are proportionally more of them where probability is higher, *or* we have weights on the samples that play the role of "number of samples". The samples could be diachronic—e.g. the firing rates of a neuron at n different times could be used as n samples from a target posterior, provided the distribution of firing rates over time matches the shape of the posterior we are interested in. Or they could be synchronic – e.g. the firing rates of different neurons in a population at a time can be treated as weighted samples from a distribution. As mentioned above, in the synchronic case, we often have an algorithm like a particle filter that starts with quite random unrepresentative samples, and then gradually refines them into a more representative set, using a "testing and resampling" procedure.

Although our samples give us a fairly literal representation of the posterior, we have to be a little bit careful about the sense in which there is a "structure preserving mapping" here. Because we may not have samples corresponding to every hypothesis in the state space, the samples are a discrete approximation of the full posterior. This is rather like the situation when we represent a continuous physical variable using a discrete analog representation – e.g. we represent height by piling uniformly shaped bricks one on top of another. This is recognizably a kind of structure preserving mapping, but not one with a one–one mapping from stimulus feature space to representational space. There are also subtleties about what probabilities are represented for hypotheses for which we happen to have no samples. As mentioned above, one would probably not assign these zero probability, but calculate their probability as an interpolation of sample densities as nearby values – that is, we assume that the represented probability function is fairly smooth and regular.

Whether the system really "makes assumptions" of this kind depends again more literally on how the samples are actually *used*. An important preliminary point about

[7] That said, of course if we want to decode likelihood directly (e.g. to figure out maximum likelihood), we will have to perform the linear operation that converts the first stage function into the log likelihood function, and then e.g. extract the maximum from that. So if "relevant use" includes explicitly extracting and representing likelihoods or probabilities, this further decoding process will have to occur.

use/exploitation is that a temporal sequence of samples can only be used by the system if there is an appropriate integration mechanism that counts or otherwise integrates information about the distribution of the samples. This synchronically integrated version of the samples is used more directly to guide downstream processing. It is therefore a better candidate for a state that is exploited as a representation, and a more interesting target for the kind of theoretical questions we are asking here—we therefore set the diachronic sampling case aside (see Lee (2014)) for an analogous point about temporal representation in perception).

Sampling-based codes can be used in computations that respect probabilistic norms, much as with PPCs. The difference is that the mapping function and discretization makes them better suited for different kinds of computations than PPCs. For example, whereas multiplying and dividing probabilities is easy using PPCs (because they use log likelihoods), adding and subtracting probabilities is easier with samples (because you can just add and subtract numbers of samples). Samples are also easier to use for computing marginal probabilities (as when you compute the probability of a function of variables, like the probability that the sum of two variables sums to a certain number), and have computational advantages when it comes to the operations involved in learning (i.e. updating conditional credences) (Fiser et al. 2010).

In the preceding, we have been assuming that the "signals" in SDT, PPC and sampling models are variable physical quantities (e.g. firing rates), a crucial part of the notion of "analog representation". This is strictly speaking a claim at the implementational level, because it tells us about the character of the *vehicles* of representation. We think this is typically assumed by proponents of our models, and justifiably, because it does explanatory work. How so? Note that a "thin", purely algorithmic-level reading of each of the models is possible, that is neutral on this implementational issue (thanks to Jake Beck for pressing us on this). We can then ask what is lost explanatorily from such a thin interpretation. Consider, for example, the signal in signal detection theory. It is typically assumed to be a variable physical quantity, like the firing rate of a neuron or the speed of a windmill, of the kind suited to be an analog representation. But one could imagine instead, say, a display with digital numbers on it, and a threshold for a decision being based on the number represented on the display being above a certain number. Provided we can model the variability in the digital signal in the same way we could an analog signal, SDT still applies here. Similarly, one could have a probabilistic population code model on which the elements in a code are digital signals rather than analog signals like neural firing rates. What would be lost from leaving open the possibility of such an implementation?

There are different answers for the different models. In the case of probabilistic population codes, as we mentioned, a selling point of the scheme is that many operations on probability distributions are reduced to linear operations on firing rates (e.g. adding them or subtracting them), which are known to be easily implemented in the brain, and which are a classic example of analog computations. So here we have a crucial bottom-up motivation for the view which appeals to the implementation level. Similar points can be made about sampling models: in so far as *number of samples* is the variable quantity used to represent probability there is clearly not an alternative implementation that does not use this variable quantity, and still is recognizably a sampling model. With the variation on which weighted samples are used, one could in theory have "digital weights" that are not variable physical quantities, but the latter are probably motivated by considerations of neural implementation of the relevant computations (e.g. adding weights) in a way similar to the case of PPC.

The more tricky case for us is signal detection theory. If the only use of the signal is to make a binary decision using a threshold (a very thin use), then there's no reason why this computation couldn't be done using a digital representation of signal magnitude, rather than a physical magnitude directly. So in a sense, SDT *qua* theory could be said to be neutral on whether the signal is analog or digital ("it's just a bunch of math!"). That said, in realistic settings in the human brain, there are typically features of the model that favor an analog interpretation—for example, the assumption of normal noise might be best explained by an analog neural signal. Furthermore, once we get to the more rich kinds of computational use of the signal we mentioned earlier, there could be computational motivations for regarding the signal as analog (e.g. if the threshold is computed using a prior). We think it's probably best here to just distinguish between a thin SDT model that is implementation neutral and a thick version which assumes an analog signal: our comments above concern the thick version of the view.

Now, as mentioned, synchronic sampling codes, PPCs, and SDT-style signals could also all get representational status by being consumed at a personal-level. A particularly vivid way in which this might happen is if they contribute to the phenomenology of experience, so that uncertainty is in some way represented in the content of experience. But what would it take for that to happen, and what reasons do we have to suppose that it ever does happen? That is the subject of the final part of the paper.

5 Uncertainty in Conscious Experience

Many theorists have the intuition that consciousness commits to a single definitive interpretation of the world, with no representation of uncertainty (Block 2018; Clark 2018). Others have argued in favor of uncertainty in experience, holding a view on which experience can assign degrees of confidence to propositions in a way analogous to degrees of belief (Morrison 2016; Morrison Manuscript) and Munton (2016)). This is the version of "experiential uncertainty" we will consider here, although probability could in theory show up in experience in other ways (for example, mean and variance could be explicitly represented).

A distinctive feature of this debate is that both parties agree that whether there are probabilistic representations in subpersonal processing does not immediately settle the question of whether conscious experience is itself probabilistic. For example, as we discuss in Section 6, there might be functional constraints on consciousness that make it more efficient to settle on a single interpretation. Indeed, some conceive of this point as a puzzle: if perceptual processing is probabilistic, why is consciousness not probabilistic? (Block 2018).

A second distinctive feature of this debate is that it can be hard to know what we are looking for—what would experiential confidence feel like, or what would it consist in? Our goal is to answer this question rather than the question of whether experiential confidence exists (in line with our goal earlier in the paper). Nonetheless, for this purpose it will be useful to consider two arguments that it exists, the basing argument (Morrison 2016) and the pipeline argument (Morrison manuscript). This is because these arguments can be seen as recommending we interpret certain experiences as involving experiential credences—a recommendation we will reject given our positive account. We are open to the possibility that experiential credences are at least *possible* however, as we discuss in the final section.

In line with our proposal in the previous section, our view is that experiential confidence should be understood as a variable phenomenal property (i.e. a feature that contributes to how a perceptual experience feels subjectively) that exclusively and systematically plays the role of indicating a level of confidence, in the manner of an analog representation. Importantly, we remain non-committal as to what this phenomenal feature would feel like. For example, we are not committed to the view that the phenomenal feature would feel "confidence-like", or that it would be a feeling of certainty that a certain environmental element is present. We think the subjective consideration that we do not seem to find a feeling of certainty (or uncertainty) in perception is a red herring. We think that perceptual experience would be probabilistic in an appropriately robust sense if there were a probabilistic state in perceptual processing that met the further functional conditions required for being a phenomenal state (e.g. that it is globally broadcast). It need not be a paradigm experience—e.g. a state that the subject very confidently would describe as a probabilistic experience—but rather one whose existence can only be confirmed by theoretical inference from our best theory of consciousness. Indeed, if probabilistic experiences are not paradigm in this way, that would explain how there can be uncertainty about whether they exist. Our view is that since the most empirically plausible models of probabilistic states postulate analog probabilistic representations, the most promising version of experiential confidence to investigate is one on which there are also such analog representations systematically and exclusively indicating probabilities. That having been said, as we will see, there are also weaker senses in which we can "experience probability", but they are much less interesting, and will be acceptable to most theorists.

On a strong reading of our proposal, it amounts to the requirement that there are phenomenal features of experience that are *intrinsically* probabilistic states, or probabilistic states that are *fully grounded* in phenomenology – a kind of "phenomenal content". A weaker view would be that there are phenomenal features that play the role of probabilistic states (e.g. they are the vehicles of probabilistic states), but they are partly probabilistic states in virtue of their extrinsic features like their functional role. Let's call these *strongly phenomenal* and *weakly phenomenal* experiential confidences.

The trouble with strongly phenomenal confidence is that it's unclear whether *any* of the more important kinds of content that experience has are strongly phenomenal (see Papineau (2021). For example, one of us doubts that even things like spatial and temporal contents are fully grounded in phenomenology. We don't want to build any such strongly intentionalist view of phenomenal properties into the issue we are considering. The trouble with weakly phenomenal confidences is that they raise the question of what exactly counts as phenomenology playing a confidence-like role. It is at this point that we propose our positive view.

The contrast between our view and a weaker view of experiential credence can be illustrated by considering some of the intuitive examples that are sometimes used to motivate experiential credence. These examples tend to involve either uncertainty in a categorization of an object – for example, recognizing a friend – or a comparison of features, for example, "line A is longer than line B".

(1) Is that Kendrick I can see over there? (categorization)
(2) Is this musical interval a perfect fifth? (categorization)
(3) Is line A or line B longer? (comparison)

In these examples, if I form an unconfident judgment (e.g. being 50% confident that this is Kendrick), this judgment is in some sense *based on experience*; it's not like we have a kind of "blindsight" here where the judgment feels like direct intuition about the world that just comes to us.

This has led some theorists to argue that we must be *endorsing* our experience in these cases, in the strong sense that there is e.g. an experiential confidence of 50% that this is Kendrick. However, prima facie, these examples can all be naturally interpreted as involving a decision or categorization process that operates on an experience of features which does *not* itself involve confidences. In the "Is it Kendrick?" case, for example, we can imagine a face (or person) recognition neural net that takes as input a representation of lower-level features of a person – e.g. spatial features of the face, things like hair color, manner of walking, etc. – and outputs a certain confidence that a particular person is present. If at a distance there are only fairly sparse features represented, this may result in only a low or middling confidence. In this case, the lower-level features could be phenomenally experienced, and *feel to the subject like the rational basis for their judgment* – but that does not require that there is a separate element of experience that is a systematically varying confidence-about-Kendrick phenomenal parameter.[8]

A similar model works well for comparison examples. For example, two spatial distance experiences of lines A and B might be grounded in two representations that feed into a subtraction process generating a signal used to decide whether A or B is longer in a way that involves probability.. There is no need to suppose here that this signal is itself perceptually phenomenally present, as opposed to the representations underpinning the individual distance experiences (see Beck 2020 for a similar take).

It's important to distinguish different reasons why categorization or comparison might lead to lack of certainty. Examples like face-recognition suggest an optimized

[8] Notice also that there is no need for background beliefs or background information explicitly represented in cognition to be playing a role here, as Munton (2016) and Morrison (2016) suggest.. The only "background assumptions" are those that the design of the recognition-net implicitly relies on, given the setting of its connection strengths.

or quasi-rational response to a lack of information about features relevant for classifying. But there are at least three other cases. One is when the stimulus is a borderline case of a category. The second is when there is noise in the categorization or comparison process itself. For example, if we have two feature representations that might lead an ideal Bayesian interpreter to judge that A is greater than B with 55% confidence, noise in the comparison process itself could lead to lower confidence in a real observer. The third case, which we think is particularly interesting and overlooked, is when the system has not fully learned the relevant categorization skill. Musical interval perception and categorization is a great example of this. Categorizing musical intervals is a fairly difficult skill to acquire that requires a lot of practice. This is the case, even when we have completely clear perception of the pitches of individual notes, so that their interval is a trivial consequence of how they are perceived. For example, I might clearly perceive a middle C, and immediately subsequently (or even concurrently) perceive the G one perfect fifth above it, but be uncertain whether the interval I hear is a perfect fourth or a perfect fifth. To sum up, in addition to lack of information, three reasons for cognitive perceptual uncertainty are borderline cases, decision process noise, and underlearning.

Again, these are not cases where the cognitive uncertainty feels to the subject like it is blind, coming from nowhere, and not based on experience. On the contrary, one would typically be attending to one's experience and feeling like the judgment is "based on experience" in these cases. Nonetheless, one can question the inference to the claim that confidence itself is phenomenally present. The musical interval example is perhaps the most compelling in this respect. One has a completely clear perception of the notes, and even (perhaps) perceives that *as* a perfect fifth, in a non-cognitive sense. Still, the conversion process required for categorization can be highly imperfect in a way that leads to great cognitive uncertainty about what is perceived.

If such a "basing" scenario is not sufficient for experiential credences, what is? As with theorizing subpersonal subjective probabilities, we think that there are more or less robust ways of making sense of phenomenology playing a confidence-like role. The examples we just considered suggest that a thin way of construing this claim, on which it is sufficient that the experience be the input into a computational process that outputs a judgment-level confidence or other graded behavior, is too weak to be interesting. To be sure, this gives us a fairly strong kind of *access-consciousness* (the delivery of information from perceptual to cognitive systems involved in reasoning, planning and memory consolidation) for probabilistic information, where this access involves a cognitive state that is based on a phenomenal state. But this doesn't seem to be sufficient for the probabilistic information to be *explicitly represented* in the experience—we have access-consciousness without phenomenal consciousness. But what does that mean—what more would that take?

Following extant literature in cognitive neuroscience, Morrison (manuscript p. 12) holds that phenomenal contents are represented explicitly when they are "useable in inference". By this, what is meant is that a representation contains probabilistic information in a form that could in principle be decoded in a relatively small number of computational steps with little or no extrinsic information. This accords with our idea that probabilistic information must be exploited by the system.

However, applied to experience, this criterion alone doesn't take us beyond the kind of access-consciousness we find in the face-recognition case (interpreting it in our sceptical way). This would make the claim that experiential credences exist both very uncontroversial, but also much less interesting. This is why we propose that experiential confidence must involve phenomenal features of experience that are analog representations—they *systematically and exclusively* play the role of confidences, in addition to playing the kind of computational role Morrison describes.

To further illustrate the point, consider a situation where an organism learns that the colors of certain boxes involve different chances that a reward is inside the box. This seems to be a case where, in context, the color representations contain probabilistic information in a form that can be decoded in a relatively small number of computational steps to deliver cognitive level credences. These color representations would then count as experiential credences in the sense described above, even though intuitively they are just color experiences. One reason this intuition seems right is that color experiences do not indicate confidences in *other* contexts: so they aren't exclusively playing a confidence role; rather they are playing the role of presenting a feature of the object which can then be used to form credences about the presence of a reward. Another is that the colors need not be systematically related to the resulting confidences—similarity and difference in confidences need not map directly onto color similarity in this case.[9]

Faced with an alleged case of experiential credence then, we think that the right question to ask is whether a stronger condition like this is met, or whether we are really dealing with something more like the colored box case. Is the subject really endorsing a confidence that is explicitly manifested in the character of experience, or merely *basing* their judgment on experience without *endorsing* the experience (as we might put it)?

Morrison, in more recent work, seems to agree with us. He gives an argument for experiential credence that does not conflate access and phenomenal senses of "consciousness". Morrison looks at cases where probabilistic information is arguably represented by subpersonal processing prior to or coincident with the representations that ground experience, and where that same probabilistic information is made available to cognition downstream. His argument is that a plausible inference-to-the-best-explanation is that this content is fed through experience into cognition. We will refer to this argument as the "pipeline" argument.

On a weak reading of this argument, "going through the pipeline" merely requires that probabilistic information implicit in subpersonal states be made explicit by the processing leading to cognition, and therefore access conscious. This would be vulnerable to the objections we just made. We think the argument is more compelling

[9] Note that above when we discussed PPC representations, we saw that although a "change of basis" was needed to explicitly decode log likelihoods, these representations were *already* useable in probabilistic inference without the need for prior decoding. This also suggests we should be open to the possibility that what is phenomenally present corresponds not directly to probabilities, but to a systematically related quantity (log liklihoods would be another example related to PPCs). One should also note that in the box-color example, the states themselves are also not phenomenally represented – there is no separate phenomenal element that means "reward" or some specific kind of reward.

if we interpret the upstream states as systematically and exclusively representing probability in something like the way that we described in the previous section. If confidence is explicitly represented in this strong sense throughout the stages of processing associated with conscious experience, that does indeed give us some prima facie reason to think that this is manifest in phenomenology in some relevant sense.

However, what we find less compelling are the actual cases where this is supposed to occur (see also Siegel 2020). The cases considered above such as face recognition are not those that Morrison appeals to; his interpretation of these is only indirectly supported by the argument. One might think that this is because the process leading to, e.g., an uncertain face-recognition judgment is clearly downstream of experience (it takes perceived low-level features as input), and so the argument doesn't apply—we return to this point below.

Let's consider a case where the pipeline argument is supposed to work. Morrison thinks that certain inter-modal integrations, where the perceptual system combines information from different sensory sources about a single target feature, are going to have the right features. In some of these cases, the perceptual system appears to be sensitive to the amount of noise in different channels, and uses it to weight the channels in a rational way (relatively noisier channels get proportionally less weight). Morrison argues that these noise representations are fed through the pipeline into cognition – so we experience noise levels in some sense. An example is a visual-vestibular integration experiment by Fetsch et al. (2012), where a subject is on a moveable platform looking at target stimulus of moving dots, and has to determine which way on average the dots are moving based on (possibly conflicting) visual and vestibular cues. The experimenters manipulate the amount of visual noise by manipulating the stimulus – e.g. they change the level of coherence of the dot motion (as Morrison notes, they could have also added vestibular noise by, e.g. adding vibration). The idea is that the process integrating these signals based on their noise levels is prior to consciousness, and so there is an explicitly represented confidence associated with the feature *direction of dot motion* that is "fed through the pipeline" in the relevant sense, and so a strong candidate for being phenomenally present.

However, the problem with this is that the subject experiences low level features—dot coherence, or level of vibration—that systematically correlate with the relevant noise parameters. Furthermore, the system could use these feature representations as a proxy for noise levels, even if they don't primarily function to represent noise (e.g. they don't do this systematically or exclusively). So there's no good case for thinking that noise levels are explicitly experienced here (not that this was Morrison's claim). Moreover, the process leading to uncertainty about direction of dot motion, since it is based on these other (conscious) feature representations, is *downstream* of experience. And so the pipeline argument would not seem to get a grip here after all; the case is too much like the "colored box" case.

Morrison's response to this (informal communication), is to say that, for the pipeline argument to work, it is not required that the processing that leads to uncertainty being explicitly represented be prior to *all* conscious representations. It is only required that it is prior to the processing that leads to whatever experience the subject has of the relevant feature (e.g., in this case, direction of dot motion); the process may be "sandwiched" between conscious representations. So if we know that

(1) the subject experiences (in some sense) direction of dot motion
(2) there is a computation of an explicitly represented confidence about dot motion that is prior to the representational basis of (1), and
(3) this confidence is fed through to influence the subject's confidence in judging direction of dot motion

then we have a plausible version of the pipeline argument that is, the experience in (1) involves an experiential confidence.

We certainly agree that (1), (2) and (3) would give a plausible argument for experiential confidence. The trouble is that it's now harder to be sure that (1) and (2) are both true in the relevant case. In particular, it's not obvious that (2) is true; for example, it could be that the subject does not experience dot motion at all, and merely judges it, or it could be that the subject has separate modality specific experiences of dot motion that are merely rationally combined into a judgment by a process of inter-modal integration that occurs downstream of experience in the relevant sense. We don't positively endorse these alternative interpretations, but merely want to suggest that more work may be required to rule them out.

It's worth commenting here that this weak version of the pipeline argument could actually be used in the face-recognition and categorization cases above. For example, even if experienced low-level features are an input into face-recognition, face-confidence could still "go through the pipeline" provided we (1) have an experience "as of Kendrick" and (2) computation of Kendrick-confidence is prior to this experience (if not all experience). What the debate might turn on, at that point, is whether we have any perceptual face-recognition phenomenology *at all* in addition to low-level phenomenology, and whether this phenomenology is a result of the same process that leads to unconfident Kendrick judgments. We think neither of these claims is obviously correct, but also not obviously incorrect either. Morrison's pipeline argument therefore is at least still promising, once understood correctly.

Finally, we think it is instructive to consider the case of blurry vision, which is a strong *prima face* candidate for involving a kind of perceptual confidence. Blurry vision is one of a class of cases where a neural representation is ambiguous between a probabilistic reading and a straightforward feature-representation reading involving feature-strengths (i.e. representations of contrast)—we call this strength/confidence ambiguity (Denison (2017) and Denison et al. (2020) both briefly float a similar idea). How the ambiguity is resolved depends on how the representation is used; but this suggests that if it grounds an experience like an experience of blur, this experience does not exclusively function to represent confidence, and therefore is not a paradigm experiential confidence in our sense.

Let us elaborate. Compare the case where you very clearly see the location of a thin line on a page, versus the case where you only have a quite blurry perception of its location. Blurry perception does not prima facie need to be understood in terms of confidence at all. It can simply be understood as a case where you have less information about the stimulus (technically, you perceive its lower frequency components but not its higher-frequency components – it is just like muffled auditory perception in this way). At this point there is already an ambiguity, depending on whether the stimulus is interpreted as simply objectively lacking these components (i.e. the line is printed blurrily on the page), or instead you interpret the lack of high frequency perceptions as indicating that there is more to be learned (e.g. by looking closer, or putting your glasses on). None of this yet involves degrees of uncertainty being represented in any robust sense though. Still, it's at least conceivable that blurry perception could also represent uncertainty. How so?

Suppose the blurry percept is partly realized by a neural population code. Crudely, suppose we are considering a small patch where the vertical line is, and we have a bank of neurons whose receptive fields vary horizontally across the patch. We get a very spread out firing-rate distribution with blurry perception, and a highly peaked one with sharp perception. Now, on one way of interpreting these neurons, they represent stimulus strength – that is, e.g. *contrast* in the preferred location in their receptive field. But since each is also driven to some extent by input from neighboring locations, they can also be given an interpretation in terms of probability of high-contrast, sharp line at a certain location. Compare the windmill example – in addition to being sensitive to wind direction, windmills are also sensitive to wind-strength. In our above example, we assumed that this was held constant, but it need not be. Thus our bank of windmills could be ambiguous between a non-probabilistic representation of windspeeds in different directions, vs a probabilistic representation of what the predominant wind direction *might* be.

Clearly, what kind of representation we have here depends on how the population code / windmill bank is *used*. If it is used flexibly for both strength and confidence purposes, it could even simultaneously be both kinds of representation.

How does this affect the debate about perceptual confidence? Suppose we are thinking about phenomenally conscious blurry vision. If my experience is used in the right way, am I consciously perceiving a probability distribution over the possible locations of the line or edge?

Notice that here there is no perceived property of the world that indicates that uncertainty is appropriate. Instead, it is a *lack* of perception of more fine-grained detail, detail which might antecedently be expected to exist, that is cue that perception is noisy. Moreover, the level of noise might be easily read out from the spread of the population code—and if this is phenomenally present, there could be a rather literal way in which noise levels are correlated with a feature of experience, which could thereby be thought of as an "analog representation" of noise.

Nonetheless, there are important caveats about embracing this as "experiential credence" in some sense. First, as Vance (2021) emphasizes, if what is represented is noise, then this is technically a likelihood (the distribution of internal signals given a stimulus property), not a posterior; so we are not attaching posterior probabilities to external states. This would still be probabilistic content represented in experience in a broad sense, however, that would be directly relevant to forming credences (e.g. if vision is unreliable, I should be less sure about judgments based on it). Here the issues mirror those that we discussed above concerning the interpretation of population codes.

Second, the core blurriness phenomenology here is still potentially fully explained by a non probabilistic representation that can be present without any probabilistic representation—namely, a representation of the lower-frequency spatial components of the spatial layout. The probabilistic gloss on this is much like the gloss on it by the system *as* perceptually-caused-blur as opposed to leaving it open whether the perceived blur is because the world is objectively blurry (as in looking at a blurry photograph), or because of a perceptual inadequacy (we probably also sometimes perceive things *as* objectively blurry—the photo example is a nice case). The tricky issue here is whether this "further interpretation" is enough to say that we have a different kind of experience in an important sense (e.g. does it make a further phenomenal difference?), despite a core phenomenal similarity. Similarly, what would count as using the experience *qua* representation of noise, such that the perception of blur counts as, or is a component of, an experience of noise? What difference is there between saying this and saying that the experience itself is simply a blurry spatial experience, and that the blur is a *cue* to what the noise is?

It's worth mentioning here the analogous case of auditory muffling, which is perhaps more intuitive. Prima facie, muffled auditory experiences just present the low-frequency components of the sound, perhaps leaving it open whether there are also unperceived high-frequency components. If an interpretation of the source of muffling is made by the system, (e.g. it's because I'm listening from behind a wall, or I'm listening on poor speakers, or I have wax in my ears), this might affect the experience in some sense (e.g. if I think the sound itself is not objectively muffled, I might have a "sense of absence"), even though it's also true that there is some core phenomenological sameness in the muffled phenomenology that is preserved. If we now add on the table the hypothesis that *noise* (a statistical property) is sometimes represented by the muffledness of auditory experience, a similar point applies. Muffled hearing is certainly a source of uncertainty—e.g. the unclarity about consonant timing and frequency distribution could make speech far more ambiguous. But what would count as the experience itself representing these kinds of uncertainty or noise, as opposed to this being merely a downstream interpretation?

Much cleaner would be a case where we can vary the noise without also changing other kinds of contents (e.g. the high-frequency spatial or auditory components). Compare hearing the same story from a trustworthy source vs an untrustworthy source. Certainly in theory one could have two channels that represented exactly the same contents, but one resulted from a much noisier process, and so is more unreliable. If the system could be sensitive to this fact about itself, and this be fed through into perceptual experience, so that a distinctive phenomenal feature indicates noise level independently of features like muffledness or blurriness, that would be a far more clear-cut example of experiential probability. As it is, we have phenomenology that primarily has a non-probabilistic role, and which might sometimes have a probabilistic interpretation. We are not aware of any cleaner examples, and we think it would be unsurprising if noise is only accessible to the system through correlated feature representations. This discussion therefore suggests that we do not yet have any clear examples of probabilistic representation in experience in more than a quite limited sense.

6 Restrictions on Consciousness

Finally, let's ask: what might explain restrictions on probabilistic contents in perception? Let's first distinguish three views of what restrictions exist. On the Definitive view, perceptual experience never has probabilistic contents, always presenting a definitive interpretation of the world. On the Restrictive view, certain kinds of probabilistic content can show up in experience, but others can't (for example, perhaps we can never experience a bimodal distribution, or we can't experience higher-order statistics like variance). On the liberal view, there are no restrictions on the varieties of probabilistic experience. Restrictive views can be read with different modal strength. On an essentialist view, restrictions hold necessarily as a result of the nature of experience itself; a state simply could not be a conscious perception and also have certain kinds of probabilistic content. On a weaker contingentist position, the restrictions do not hold necessarily, but simply are the result of certain contingent features of our cognitive architecture.

Suppose we hold a representational view of consciousness, in the sense that we think that conscious states are mental representations meeting certain further conditions that make them conscious – for example, they are globally broadcast, or they are targeted by higher-order thoughts (on many theories, this means they are representations that are *consumed* in a certain way). In principle, these further conditions could entail that essentialism is true. For example, Andy Clark (2018) proposes that conscious perception is definitive because its role is to guide action, and motor systems need a definitive take on the world in order to act on it. A strong version of this view would be a kind of essentialism on which being action-oriented in the relevant sense is essential to consciousness, and necessarily implies definitiveness. However, we suspect that this reading of Clark's view is implausible, and that the same will tend to be true of other essentialist positions, if they are based on plausible views of consciousness. An appropriate test is whether we can coherently imagine a robot or alien that has a cognitive architecture similar to ours, including having representational states that play a similar functional role to our conscious states, such as being action-oriented, but which uses a wide range of probabilistic states for these purposes. Take Clark's "action-guidance" proposal: there is simply no reason why, at least in principle, action can't be guided by probabilistic information rather than a single definitive proposition about the world. This is already obvious from the way that our cognitive states can guide action without being definitive. When one thinks about other well-known proposals about the essential functional role of consciousness (e.g. global broadcast), the same point tends to apply. Of course, we don't know what the correct theory of consciousness is, and so it could turn out to contain a relevant constraint; but we see no independent reason to believe this given the current state of play.

All this is not to say that there might not be contingent features of our human cognitive architecture that constrain the kinds of contents we can experience. For example, perhaps Clark's action-guidance constraint is plausible given the (contingent) computational cost of keeping uncertainty in play. In general, what kind of contingent constraints should we consider? We would distinguish input end constraints – constraints that come from the way contents are selected for consciousness – and functional constraints on the way information is used once it is conscious. On the input end, one hypothesis is simply that there do not exist probabilistic representations in a robust enough sense in perceptual processing for there to be candidates to become conscious (Denison et al. 2020 suggest this position). Another input end constraint to consider is that there is a low bandwidth bottleneck for content to become conscious, and that the system deals with this by insisting on a definitive interpretation. On the function end, it may be that some or all of the uses that conscious information is put to are more easily accomplished with a definitive content, in a way that outweighs the cost of throwing away potentially useful probabilistic information.

For example, Dehaene (2014, ch.3) suggests that

> "The function of consciousness may be to simplify perception by drafting a summary of the current environment before voicing it out loud, in a coherent manner, to all other areas involved in memory, decision, and action. In order to be useful, the brain's conscious brief must be stable and integrative. During a nationwide crisis, it would be pointless for the FBI to send the president thousands of successive messages, each holding a little bit of truth, and let him figure it out for himself. Similarly, the brain cannot stick to a low-level flux of incoming data: it must assemble the pieces into a coherent story. Like a presidential brief, the brain's conscious summary must contain an interpretation of the environment written in a "language of thought" that is abstract enough to interface with the mechanisms of intention and decision making."

Here it is worth explicitly comparing Clark's "consciousness is for action" explanation with a broader kind of view that appeals to a range of different uses of conscious information. On a view like Dehaene's, conscious information is broadcast into a global workspace, where it can be used by a variety of different cognitive processes, not just action control. It could be that it is this need for a "one size fits all post-perceptual uses" kind of representation which forces conscious content into a definitive form.

Of course, one might reject the definitive view. Perhaps *some* uncertainty is phenomenally conscious. It is important to note, however, that any *restrictions* on probabilistic content raise a similar explanatory challenge to the one facing the definitive view, and similar kinds of considerations may be appealed to. For example, keeping estimates of noise in play (e.g. in blurry vision) may be consistent with "one size fits all" even if keeping many hypotheses with separate probabilities in play is not.

Our tentative rejection of essentialism partly depended on regarding conscious experiences as mental representations. Now, we suspect that the intuitive appeal of the anti-experiential-credence view is partly related to the intuitive appeal of non-representational views of consciousness like naïve realism. If experience is just a direct confrontation with concrete features of the environment, then there is no room for hedging or uncertainty in it. Conversely, if we imagine a being with a more thoroughly probabilistic kind of perception than ours, such as our robots or aliens above, then surely they would be less tempted to regard experience as direct acquaintance with the environment, as opposed to a representational guide.

Having said this, we do think that one possibility to take seriously is that conscious perception is contingently configured so that its phenomenal features always seem to present concrete external features of stimuli (it is "transparent"). This would have the result that, even if there were features of experience that played the functional role of representing uncertainty, we would not experience them *as such*, we would experience them as just presenting the stimulus as configured a certain way (e.g. as blurry). This is perhaps one promising way to explain the intuitive resistance or puzzlement many have to the very idea of experiential credences.

7 Conclusion

To conclude, we looked at three models of subpersonal probabilistic representations, and the case of probabilistic representation in conscious experience. Without assessing whether the models accurately portray sub-personal processing, we gave an account of the sense in which they postulate probabilistic representations: such representations are analog representations of probability in the sense explained above. This account has the benefit that it transfers in a very natural way to the case of experiential credence, giving us a robust and plausible sense of "experiential credence", on which it is an interesting and potentially controversial claim that such credences exist. Our view is that although there are no clear cut cases of experiential credence (in this sense) that have been demonstrated to exist in humans, it is still plausible that such credences are at least possible. This is because the relevant restrictions on what can be represented in conscious experience are most plausibly regarded as contingent; it is a quirk of the way that consciousness is configured in us that we do not encounter the world in experience in a richly probabilistic way.

Declarations

Conflict of Interest None.

References

Beck, J. 2019. Perception is analog: The argument from Weber's law. *The Journal of Philosophy* 116 (6): 319–349.

Beck, J. 2020. On perceptual confidence and "Completely Trusting Your Experience." *Analytic Philosophy* 61 (2): 174–188. https://doi.org/10.1111/phib.1215.

Block, N. 2018. If perception is probabilistic, why does it not seem probabilistic? *Philosophical Transactions of the Royal Society B: Biological Sciences* 373: 20170341. https://doi.org/10.1098/rstb.2017.0341.

Bowers, J.S., and C.J. Davis. 2012. Bayesian just-so stories in psychology and neuroscience. *Psychological Bulletin* 138 (3): 389.

Clark, A. 2018. Beyond the'Bayesian blur': Predictive processing and the nature of subjective experience. *Journal of Consciousness Studies* 25 (3–4): 71–87.
Colombo, M. 2013. Constitutive relevance and the personal/subpersonal distinction. *Philosophical Psychology* 26 (4): 547–570.
Davies, M. 1995. Two notions of implicit rules. *Philosophical Perspectives* 9: 153–183.
de Finetti, B. (1937/1964). Foresight: Its logical laws, its subjective sources. In H. E. Kyburg Jr & H. Smokler (Eds.), Studies in subjective probability (pp. 94–158). Originally published as "La prévision; ses lois logiques, ses sources subjectives" in Annales de l'Institut Henri Poincaré, Volume 7, 1–68. New York: Wiley.
Dehaene, S. 2014. *Consciousness and the brain: Deciphering how the brain codes our thoughts*. Penguin.
Denison, R. 2017. Precision, not confidence, describes the uncertainty of perceptual experience: Comment on John Morrison's "Perceptual Confidence." *Analytic Philosophy* 58 (1): 58–70.
Denison, R., N. Block, and J. Samaha. 2020. What do models of visual perception tell us about visual phenomenology? In *Neuroscience and Philosophy*, eds. F. De Brigard and W. Sinnot-Armstrong. Cambridge: MIT Press.
Drayson, Z. 2012. The uses and abuses of the personal/subpersonal distinction. *Philosophical Perspectives* 26 (1): 1–18. https://doi.org/10.1111/phpe.12014.
Dretske, F. 1997. *Naturalizing the Mind*. MIT Press.
Fetsch, C.R., A. Pouget, G.C. Deangelis, and D.E. Angelaki. 2012. Neural correlates of reliability based cue weighting during multisensory integration. *Nature Neuroscience* 15 (1): 146–154. https://doi.org/10.1038/nn.2983.
Fiser, J., P. Berkes, G. Orbán, and M. Lengyel. 2010. Statistically optimal perception and learning: From behavior to neural representations. *Trends in Cognitive Sciences* 14: 119–130.
Gescheider, G.A. 1976. *Psychophysics: the fundamentals*. London: Psychology Press.
Gładziejewski, P. 2016. Predictive coding and representationalism. *Synthese* 193: 559–582. https://doi.org/10.1007/s11229-015-0762-9.
Godfrey-Smith, P. 2012. Review of Brian Skyrms' Signals. *Mind* 120: 1288–1297.
Green, D.M., and J.A. Swets. 1966. *Signal detection theory and psychophysics*, vol. 1. New York: Wiley.
Gross, S. 2020. Probabilistic Representations in Perception; are there any and what would they be? *Mind and Language* 35(3): 377–389
Jazayeri, M., and J.A. Movshon. 2006. Optimal representation of sensory information by neural populations. *Nature Neuroscience* 9 (5): 690–696.
Hoyer, P.O., and A. Hyvärinen. 2003. Interpreting neural response variability as Monte Carlo sampling of the posterior. *Advances in Neural Information Processing Systems* 15: 293–300.
Icard, T. 2016. Subjective probability as sampling propensity. *Review of Philosophy and Psychology* 7 (4): 863–903.
Kiefer, A., and J. Hohwy. 2018. Content and misrepresentation in hierarchical generative models. *Synthese* 195 (6): 2387–2415. https://doi.org/10.1007/s11229-017-1435-7.
Kirchhoff, M.D., and I. Robertson. 2018. Enactivism and predictive processing: A non-representational view. *Philosophical Explorations* 21 (2): 264–281. https://doi.org/10.1080/13869795.2018.1477983.
Lee, G. 2014. Temporal experience and the temporal structure of experience. *Philosophers Imprint* 14(3): 1–21.
Lee, T.S., and D. Mumford. 2003. Hierarchical Bayesian inference in the visual cortex. *Journal of the Optical Society of America A* 20 (7): 1434–1448.
Ma, W.J., J.M. Beck, P.E. Latham, and A. Pouget. 2006. Bayesian inference with probabilistic population codes. *Nature Neuroscience* 9 (11): 1432–1438. https://doi.org/10.1038/nn1790.
Ma, W.J. 2010. Signal detection theory, uncertainty, and Poisson-like population codes. *Vision Research* 50 (22): 2308–2319. https://doi.org/10.1016/j.visres.2010.08.035.
Ma, W.J. 2012. Organizing probabilistic models of perception. *Trends in Cognitive Sciences* 16 (10): 511–518.
Ma, W.J., and M. Jazayeri. 2014. Neural coding of uncertainty and probability. *Annual Review of Neuroscience* 37: 205–220.
Marr, D. 1982. *Vision: A computational investigation into the human representation and processing of visual information*. Cambridge: MIT Press.
Millikan, R.G. 1989. Biosemantics. *The Journal of Philosophy* 86 (6): 281–297.
Morrison, J. 2016. Perceptual confidence. *Analytic Philosophy* 57 (1): 15–48.
Morrison, J. (manuscript) "Third-Personal Evidence for Perceptual Confidence."

Moss, S. 2018. *Probabilistic knowledge*. Oxford University Press.
Munton, J. 2016. Visual confidences and direct perceptual justification. *Philosophical Topics* 44 (2): 301–326.
Orbán, G., P. Berkes, J. Fiser, and M. Lengyel. 2016. Neural variability and sampling-based probabilistic representations in the visual cortex. *Neuron* 92 (2): 530–543. https://doi.org/10.1016/j.neuron.2016.09.038.
Orlandi, N., and G. Lee. 2018. How radical is predictive processing? In *Andy Clark and critics*, ed. M. Colombo, E. Irvine, and M. Stapleton. Oxford University Press.
Papineau, D. 2021. *The metaphysics of sensory experience*. Oxford: Oxford University Press.
Pettigrew, R., and J. Weisberg. 2019. *The open handbook of formal epistemology*. Philpapers Foundation.
Pouget, A., P. Dayan, and R. Zemel. 2000. Information processing with population codes. *Nature Reviews Neuroscience* 1 (2): 125–132.
Pouget, A., J.M. Beck, W.J. Ma, and P.E. Latham. 2013. Probabilistic brains: Knowns and unknowns. *Nature Neuroscience* 16 (9): 1170–1178.
Ramsey, W.M. 2007. *Representation reconsidered*. Cambridge University Press.
Shea, N. 2014. Neural signalling of probabilistic vectors. *Philosophy of Science* 81: 902–913.
Shea, N. 2018. *Representation in cognitive science*, 304. Oxford University Press.
Siegel, S. 2020. How can perceptual experiences explain uncertainty?. *Mind & Language* 37(2): 134–158.
Skyrms, B. 2010. *Signals: Evolution, learning, and information*. Oxford University Press.
Sturgeon, S. 2020. *The rational mind*. Oxford University Press.
Titlebaum, M. 2019. Precise Credences. In *The open handbook of formal epistemology*. Philpapers foundation.
Vance, J. 2021. Precision and perceptual clarity. *Australasian Journal of Philosophy* 99(2): 379–395.