9 What Do Models of Visual Perception Tell Us about Visual Phenomenology?

Rachel N. Denison, Ned Block, and Jason Samaha

9.1 Introduction

The science of visual perception has a long history of developing quantitative, computational models to explain and predict visual performance on a variety of tasks. These models have typically been developed to account for objective visual performance, such as observer's accuracy, reaction time, or perceptual sensitivity in discriminating different visual stimuli. Much less examined is how these models relate to the subjective appearance of a visual stimulus—that is, the observer's phenomenal character of seeing the stimulus. The goal of this chapter is to examine that second link—between models and phenomenal experience.

9.1.1 What Is Phenomenology?

By "phenomenology" or "phenomenal experience," we mean the first-person, subjective, conscious experience an observer has of a visual stimulus. We are interested here in the experience of the properties of that stimulus, for example its contrast, color, or shape. That particular visual phenomenal content is to be distinguished from the mere fact of the observer's being conscious (i.e., awake vs. in a dreamless sleep), from the totality of one's conscious experience (which includes non-visual content such as thoughts, feelings, and non-visual perceptual experiences), and from the experience of visual content unrelated to the stimulus of interest (e.g., content in the periphery of one's visual field). In short, an observer's phenomenal experience of the properties of a stimulus is a matter of what a particular stimulus looks like to them. Here, we ask whether current vision models can make predictions about this kind of subjective visual experience.

9.1.2 Key Concepts in Visual Neuroscience

First, we introduce a few key concepts in visual neuroscience that will be relevant for understanding the models. Much of what we know about the response properties of visual neurons comes from studying early sensory brain areas such as the primary visual cortex (V1) while presenting relatively simple stimuli such as oriented luminance gratings (striped patches, as in figure 9.4). The properties of V1 neurons described below have informed the models discussed in this chapter, as well as attempts to link model components to neural responses. However, the models we consider are meant to be general, and could be used to describe the encoding and decision-making mechanisms for diverse stimulus properties and sensory and decision-related brain regions.

Contrast sensitivity The contrast of a stimulus refers to its variation in luminance across space—the differences between the lightest and darkest regions of the image. Visual neurons tend to respond to luminance changes rather than absolute luminance. So, the response of a neuron typically increases with higher contrast.

Orientation selectivity Many neurons in V1, and at later stages of visual processing, are selective for the orientation of edges in an image. For example, a neuron may respond strongly to a region of an image with a vertical edge but respond very little to a horizontal edge. The response profile of a neuron to different orientations (0–180°) is called an orientation tuning curve. Many vision science experiments use grating stimuli with different orientations and contrasts because extensive physiological experiments have related the responses of visual neurons to such stimuli, providing a theoretical foundation for linking neural activity and behavior.

Response variability Neurons have ongoing spontaneous activity, and they don't respond exactly the same way to repeated presentations of a stimulus. On one trial, a neuron may respond a bit more, and on another trial a bit less to the same stimulus. This trial-to-trial response variability is sometimes called internal noise. It creates uncertainty in the mapping between a stimulus and a neural response. Such response variability of sensory neurons is considered one of the sources of variability in human behavior.

9.1.3 What Is a Model?

Here, we define a model as a compact mathematical explanation of some phenomenon. Visual perception is a complicated phenomenon. So, we make models to capture and test our understanding of how it comes about—that is, we use models to explain and to predict. First, we might like to *explain* observers' behavior in a specific visual task. Why did they report seeing the things they reported seeing? That is, what are the underlying computations or processes that produced their behavior? Second, we might like to *predict* observers' behavior on another session of the same task or on a somewhat different task. How well can we predict what they will report seeing? Predictions allow us to test a model quantitatively, compare the performance of alternative models, and identify mismatches between predictions and data, which point to how the model might be improved.

Models can be specified at different levels of abstraction (Marr, 1982), allowing different kinds of explanations and predictions, depending on one's scientific question. For example, a more abstract model—at what has been called the computational level—can be specified with variables and equations that correspond to cognitive processes but bear little connection to neurobiology. A less abstract model—at the implementational level—could involve simulating large numbers of spiking neurons. The kinds of components that make up a model are quite flexible.

Here, we are concerned with observer models (Geisler, 2011; Rahnev & Denison, 2018; Kupers, Carrasco, & Winawer, 2019), which take as input stimuli and task goals and produce as output behavior and/or neural responses (figure 9.1). The output of an observer model can therefore be compared to human behavioral and neural data for the same stimulus and task. What goes inside the model can vary, but all the models we will discuss can be thought of as having two stages: a sensory stage and a decision stage. The sensory stage encodes stimulus features. The decision stage reads out, or decodes, the responses at the sensory stage to make a perceptual decision. At the sensory stage, a fine-grained representation of the stimulus feature may exist, but at the decision stage, this representation is lost. For example, in an orientation discrimination task, the sensory stage might encode the stimulus orientation as a continuous variable, whereas the decision stage converts that graded sensory response to a binary choice about whether the orientation is clockwise or counterclockwise of a decision boundary. The decision stage is more integral to models designed to

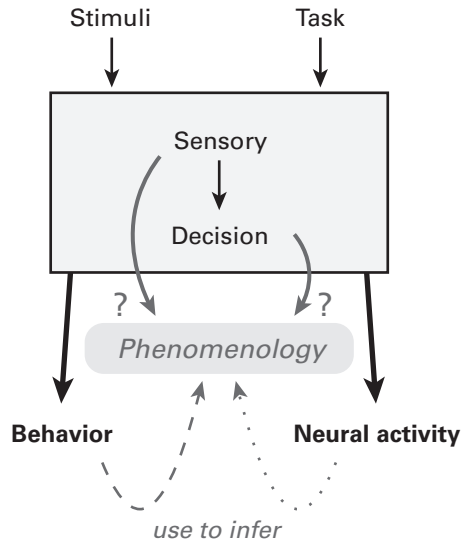


Figure 9.1
Observer models and their relation to phenomenal experience. An observer model (gray box) takes as input stimuli and task goals. The models we consider here can include sensory processing and decision stages. Traditionally, observer models are designed to output measurable quantities (shown in bold), including behavioral and/or neural responses. Here, we are interested in understanding how observer models could predict the contents of phenomenal visual experience (*phenomenology*) whose properties must be inferred from behavior and neural activity. (We use a sparser arrow to indicate inferences from neural activity because such inferences require established relations between neural activity and phenomenology in the absence of a behavioral report—an active research topic in itself.) Specifically, we ask how sensory and decision components of standard vision models could map to phenomenal experience.

produce categorical choice behavior than to models focused on describing sensory representations.

9.1.4 Models of Visual Perception and Decision Making

Here, we consider four widely used vision models: (1) signal detection theory (SDT), (2) drift diffusion models (DDM), (3) probabilistic population codes (PPC), and (4) sampling models. These models can be organized along two dimensions: (1) whether they are static or dynamic, and (2) whether they use point representations or probabilistic representations of stimulus features (table 9.1). Static models (SDT and PPC) describe sensory and decision

Table 9.1
Classification of models

	Point Representation	Probabilistic Representation
Static	SDT	PPC
Dynamic	Drift diffusion	Sampling

SDT: signal detection theory; PPC: probabilistic population codes.

variables at a single time point, or some steady state, whereas dynamic models (DDM and sampling) unfold over time (usually, for perceptual decisions, timescales of hundreds of milliseconds to seconds).[1] SDT and DDM models represent some stimulus feature as a single continuous-valued number or *point estimate*. (This term is used to mean a point representation.) This number could represent, for example, the strength of evidence for one choice versus another in a two-choice perceptual decision. In contrast, models with probabilistic representations (PPC and sampling) also contain distributional information about the uncertainty of the stimulus feature.[2] In coding a sensory feature, probabilistic models represent not only a point estimate (e.g., the distance is 5.3 meters), but also the certainty of the estimate or even the probabilities associated with all possible feature values (e.g., the distance is 5.3 meters with probability x, 5.4 meters with probability y, and so forth).

9.1.5 Modeling Phenomenal Experience

How does phenomenal experience relate to the components of these models? Here, we find a gap. These models have been developed largely to predict and explain objective behavioral performance and neural activity, not phenomenal experience. At first glance, this gap is surprising—a major aspect of visual perception is our conscious experience of the way things look. So, why hasn't visual phenomenology been a modeling focus so far? Here are three possible reasons:

First, there is good reason for a scientific strategy that does not focus on phenomenal experience. As experimenters, we can devise models that predict and explain neural and behavioral variables without probing observers' phenomenal experience per se. And whereas behavior and neural activity can be measured directly, phenomenal experience must be inferred from behavioral reports and/or neural responses (figure 9.1). As a result, ascertaining observers' phenomenal experience—for example, whether they consciously perceived a stimulus—is fraught with conundrums. Should we

use subjective measures of conscious awareness, such as asking observers if they saw something, or objective ones, such as above-chance performance? What is the relation between metacognition (e.g., observers' appraisal of their perceptual decisions) and phenomenal experience itself? Is consciousness all or nothing—or does it come in degrees? Confronting these issues is challenging and can reduce the attractiveness of phenomenal experience as a research target. While this argument reflects a real challenge in the empirical study of phenomenal experience, we believe there is still much room for progress on this front. We will return to this topic in section 9.3.2.

Second, objective and subjective measures of visual perception are often tightly coupled, which might suggest that models that explain objective reports will also straightforwardly explain subjective reports and phenomenology. For example, if an observer is asked to report whether a stimulus was one of two possibilities (an *objective report* because it has accuracy conditions with respect to the physical stimulus) and then report their confidence that their choice was correct (a *subjective report* because it only refers to the observer's internal state), correct trials will typically receive higher confidence ratings than incorrect trials. However, although objective and subjective measures are often correlated, they can dissociate—sometimes dramatically, as in blindsight (Weiskrantz, 1996; see chapter 10 of this volume, which describes such dissociations in detail). Such dissociations indicate that objective reports are not a perfect proxy for phenomenology.

Third, many perceptual phenomena—for example, rivalry and perceptual constancies—depend on unconscious processing. So, it is not surprising that models of visual perception often focus on aspects of visual processing that are not specific to visual phenomenology.

Therefore, although it's understandable why computational models of visual perception have mainly focused on capturing objective performance, there is also a need for computational models that predict the properties of phenomenal experience specifically. We think it is surprising that current standard approaches to modeling visual perception have not addressed this challenge more directly.

The goal of this chapter is therefore not to describe the precise links between current vision models and phenomenal experience, as such links have not been fully developed. Rather, to work toward bridging this gap, in section 9.2, we introduce each model according to the following structure. First, we introduce the model and describe its sensory and decision stages. Second, we discuss empirical studies, philosophical work, and theoretical

considerations that bear on how the model components could be mapped to aspects of phenomenology. Under the umbrella of phenomenology, we consider both awareness (whether some stimulus property is consciously experienced) and appearance (how exactly a stimulus looks, given that it is experienced). Finally, we summarize various options for possible mappings between model components and phenomenal experience. We view these mappings as alternative hypotheses about how model components could relate to phenomenal experience. In some cases, existing evidence can help adjudicate among these alternatives, but the matter is not yet settled. Although at this point, we cannot decide which option is best, each model provides a formal structure through which to articulate different options, which itself is a theoretical advance. In section 9.3, we discuss scientific and philosophical research directions that follow from this approach. Although the models we discuss have primarily been used to explain and predict objective performance and neural activity, they could also be applied and extended to explain and predict a variety of subjective reports.

9.1.6 Clarifications

Before we launch into the models, here are a few clarifications:

1. Are we realist or as-if (Chakravartty, 2017) about these models? We think that the computations represented by these models could be instantiated as neural processes, though the mechanics of how these computations are implemented by the brain could look very different from the model's mechanics. Note that some models (such as PPC and sampling) are explicitly neural, so the realist mapping is more direct, whereas others (SDT and DDM) are more abstract and allow different neural implementations.
2. What do we mean by a *mapping* between model components and phenomenal experience? We mean a correspondence relation with a particular phenomenal experience. One formulation of this idea is that the information represented by the model component matches the information represented in phenomenal experience. Note that when we say that a model component maps to phenomenal experience, that does not mean that the component is sufficient for consciousness. We assume these components are embedded in a functioning human brain that already meets the conditions for consciousness. Rather, we are interested in how a model component could relate to a particular phenomenal

content. Ultimately, we are interested in the neural basis of a conscious content—what philosophers call a core realizer,[3] which would require understanding the neural implementation of model components.

3. Does specifying a mapping in this way mean that conscious states are epiphenomenal—that is, they have no causal/functional properties? No. The brain activity that realizes a conscious state has causal/functional interactions within the physical system.
4. The sensory stage representations of all the models discussed here assume that the visual system has feature detectors (e.g., for orientation, motion, faces, etc.; see section 9.1.2) that respond automatically to the feature of interest. More complex or novel feature combinations (e.g., a pink elephant) may not have dedicated detectors. Such detectors may have to be learned by the brain for these models to apply to complex perceptual decisions.

9.2 Specific Models of Visual Perception and Decision Making and their Possible Mappings to Conscious Perception

9.2.1 Signal Detection Theory

SDT is perhaps the most widely applied model of perceptual decision making in psychology and neuroscience. Its popularity stems from the fact that it can elegantly explain a wide range of perceptual behaviors, can tease apart the perceptual sensitivity of an observer from response biases they may have, and provides a straightforward framework in which to model the variability of perception and behavior.

Model characterization Although the theory has been extended and built upon in many different ways, the core idea of SDT is as follows (figure 9.2). Neural detectors representing stimulus features (e.g., a horizontal line at a particular location in the visual field) respond when their preferred stimulus is present. The response magnitude indicates the amount of evidence that the particular feature is present in the environment. A detector tuned to horizontal lines at the center of gaze would respond maximally when there is a clear horizontal line present at the center of gaze, and would, on average, respond less when the image is less clear (e.g., if it were lower contrast) or when the image deviates from the detector's preference (e.g., if the image were a diagonal line or in the periphery). In this way, a single detector

presented with an image outputs a single number that reflects the amount of evidence that a particular feature is present on that trial. This quantity is sometimes called the internal response. The observer must determine: How much evidence do I require to decide that the stimulus feature was present? This amount (or threshold) is known as the criterion.

Going forward, we use the word "detector" to mean the part of the sensory stage that carries information about the stimulus feature. This could correspond to an elementary or a highly processed neural signal. For example, the neural signal corresponding to the detector's activity could be purely feedforward or could incorporate feedback. In short, the activity of the detector corresponds to the sensory representation at the computational level.

Sensory stage A core idea of SDT is that the internal response of a detector on any given trial is not determined by the stimulus properties alone but is also subject to stochastic variation (noise). This means that the same physical stimulus can produce different responses in the detector due to myriad other factors (e.g., the number of photons absorbed by the retina, the amount of neurotransmitter uptake at a synapse, the attentional state of the observer). The noise is often modeled as Gaussian with an independent mean and standard deviation. Note that talk of Gaussian distributions, in the context of SDT, refers to distributions built up over many repeated trials of an experiment. All that is represented at the sensory stage on a given trial is a single value (conceptualized as being drawn from a distribution) reflecting the amount of evidence that a particular feature is present.

Importantly, the same detector, again because of noise, will not be completely silent when its preferred stimulus feature is absent—that is, there will be another distribution from which the detector's response is drawn when the stimulus is absent. For example, many of us have confidently experienced our phone vibrating in our pocket despite no actual phone vibration. This experience can be interpreted as arising from noise in a hypothetical "vibration detector" that, due to random variability, is strong enough to be noticed. The stimulus-absent distribution is often modeled as having the same standard deviation (but lower mean) as the stimulus-present distribution, though this is a parameter that can be adjusted. Crucially, if the noise of one or both distributions is very large, or if the means of the two distributions are close, it will not be obvious which distribution the internal response on a given trial came from—that is, whether the signal was truly present or absent on that trial. In other words, the amount of

overlap between distributions determines the difficulty of the decision. This overlap can be estimated by analyzing behavioral responses to repeated presentations of stimulus-absent and stimulus-present trials and is called *d'*. As the overlap between internal responses on stimulus-present and stimulus-absent trials increases, the detection task becomes harder and *d'* decreases. Thus, *d'* reflects the sensitivity of an observer's perceptual system, which is corrupted by noise.

Decision stage Given a single sample from one of these distributions, which response should be chosen? This is the question addressed in the decision stage. SDT posits that the observer's choice is determined by whether the internal response of the detector falls above or below some criterion level. For instance, internal responses above the criterion would result in a stimulus-present response, and responses below would result in a stimulus-absent response.

A key feature of SDT is that the response criterion is independent of the sensory distributions. Thus, the amount of evidence needed to commit to a decision is theoretically separate from the sensitivity of the perceptual system. The value of *d'* could be low or high, but the observer may still require a lot of evidence to decide stimulus presence. The independence of criterion and sensitivity allows an SDT-based analysis to separate out an observer's overall propensity to give a certain response from the actual discrimination sensitivity of their perceptual system.

A point worth bearing in mind when evaluating empirical evidence is that the SDT criterion, as measured from behavior, is defined relative to the distributions themselves. So, a change to the SDT criterion can be produced by either a change in the amount of internal response needed to commit to a decision or by shifting the means of both sensory distributions while keeping the same absolute criterion (Ko & Lau, 2012; Iemi et al., 2017; Denison et al., 2018; figure 9.2B).

When modeling perceptual decisions with SDT, the criterion is understood to be a flexible decision boundary that the observer can adjust according to a number of non-perceptual factors, including the probability the stimulus is present and the rewards or costs associated with detecting or missing stimuli. We refer to this standard SDT criterion as a *decisional criterion*, and we take it as obvious that deliberate or strategic shifts in decisional criteria do not necessarily affect conscious awareness. Here, however, our aim is to map SDT to phenomenal experience. Therefore, in the

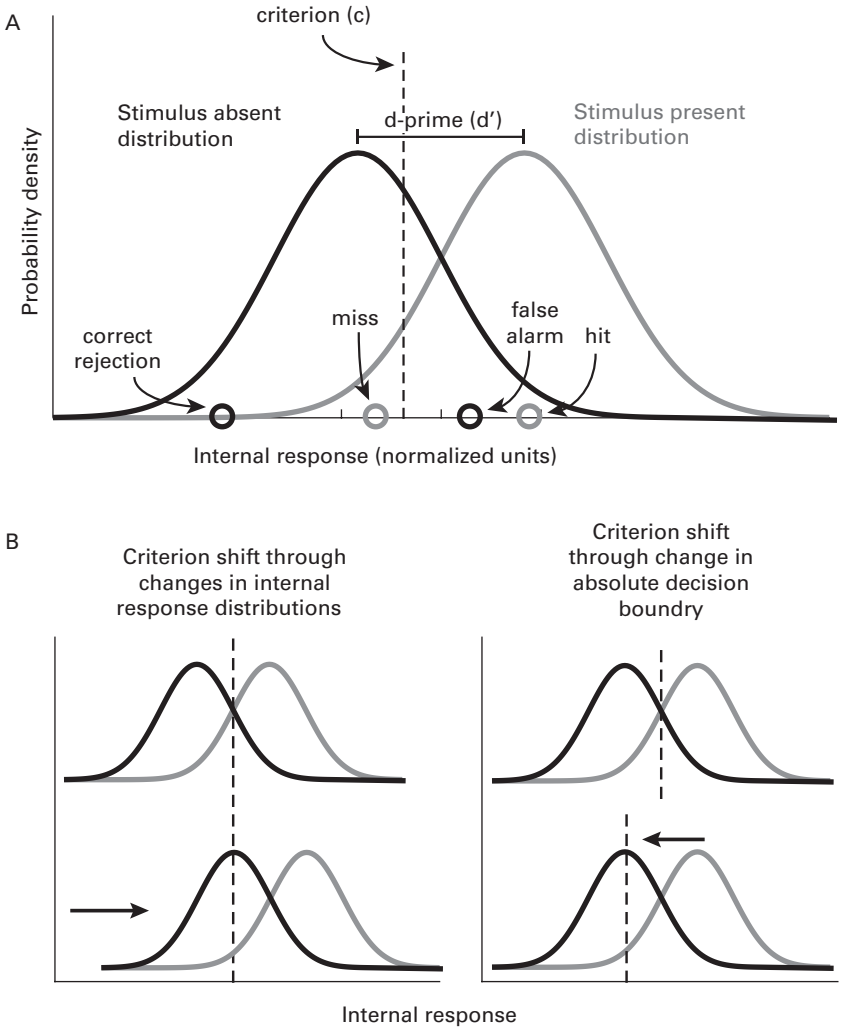


Figure 9.2

Signal detection theory (SDT) illustrated for a detection task. (A) The sensory stage of SDT is a single internal response, generated on every trial (open circles on the *x*-axis represent four example trials). When no stimulus is presented, the internal response is drawn from a Gaussian probability distribution (stimulus-absent distribution). When the stimulus is presented, the internal response is drawn from another Gaussian distribution with a higher mean (stimulus-present distribution). When the internal responses are normalized by the standard deviation of the Gaussian, the discriminability of the two distributions (d') is the difference in distribution means. The criterion represents the decision stage and is independent of d'. It defines the

next section, we pose the question of whether there is a *phenomenal criterion*, such that only when the internal response crosses this criterion is a conscious percept produced (see Wixted, 2019, for a related discussion of threshold theories). That is, if an observer could directly report their conscious experience, does conscious perception follow the sensory stage representation, or does it depend on a criterion crossing?

Where is conscious perception in SDT? Given SDT models' single-trial point representations of sensory evidence along with the two possible interpretations of the criterion (phenomenal and decisional), where in these models could conscious perception fit? Now, we detail possible mappings from SDT components to phenomenology, followed by empirical evidence that bears on various mappings.

Options for mapping model components to phenomenology

Option 1: Conscious perception follows the representation at the sensory stage. That is, phenomenal experience is reflected in the activity of the detector that responds, with variability, to the stimulus. Crossing the criterion (or not) only determines the post-perceptual decision. This option would imply that activity in the detector is always associated with a conscious experience that has a graded strength.

Option 2: Conscious perception follows the representation at the sensory stage but only occurs once the internal response surpasses the phenomenal

Figure 9.2 (continued)

magnitude of internal response needed to report that the stimulus was present. When the stimulus is absent, an internal response below the criterion (report absent) gives a correct rejection, and an internal response above the criterion (report present) gives a false alarm (open circles). When the stimulus is present, an internal response below the criterion gives a miss, and an internal response above gives a hit (open circles). (B) The SDT criterion computed from behavior is a relative measure: it indicates how much evidence is needed to make a decision relative to the two distributions. However, when relating behavior to neural activity, we may be interested in the absolute decision boundary—for example, the neural response magnitude required to report that the stimulus was present. This example illustrates a potential difficulty in inferring an absolute decision boundary from the SDT criterion in a detection task. The criterion estimated from behavior could change due to either a shift in the internal response distributions with no change in absolute decision boundary (left) or a shift in the absolute decision boundary with no change in internal response distributions (right). See color plate 1.

criterion. This option would imply that only responses that surpass the criterion produce graded, conscious strengths, whereas all responses below the criterion are equally unconscious. This situation would allow for unconscious perception when the internal response is above zero but below the criterion.

Option 3: Conscious perception depends on a phenomenal criterion and follows the (binary) representation at the decision stage. This option would imply that phenomenal experience has an all-or-none categorical character, such that an observer either does or does not experience a certain stimulus property. Information about the strength of evidence for the stimulus property (i.e., internal response magnitude) would be lacking in phenomenal experience.

Options 2 and 3 would also involve the existence of a separate decisional criterion (in addition to the phenomenal criterion) that could be flexibly adjusted to adapt behaviors to task contingencies (e.g., reward structure). Note that options 1 and 2 predict something about the quality of one's subjective percept—namely, the strength or intensity of that percept—whereas option 3 predicts only whether the stimulus was seen.

Evidence for mappings Recent experiments examining neural responses in signal detection tasks have started to shed light on what role a threshold-crossing process could play in whether a stimulus is consciously experienced. According to options 2 and 3, any signal that does not surpass the phenomenal criterion is unconscious. If we assume that confidence reports reflect conscious percepts, these options make predictions about metacognitive sensitivity, or how well confidence reports track performance accuracy. (An accurate response is one that matches the stimulus.) Specifically, options 2 and 3 predict that variability in an observer's confidence ratings for stimulus-absent responses should be unrelated to the accuracy of those responses[4] (Wixted, 2019). If a stimulus does not exceed the phenomenal criterion, then there should be no conscious access to how far away from the criterion the stimulus was. So, confidence will be unrelated to the internal response magnitude on these trials. Accuracy, on the other hand, should still relate statistically to the internal response magnitude.

This prediction has been tested by examining whether confidence or visibility judgments are higher on correct versus incorrect trials, even when the stimulus was reportedly not seen. Whereas abundant evidence makes clear that, for seen trials, the more visible a stimulus is rated, the higher confidence observers report (and the higher their metacognitive sensitivity;

Galvin et al., 2003), it is less obvious that the same is true of misses and false alarms. Koeing and Hofer (2011) used a near-threshold detection task with a rating scale that afforded a low and high confidence stimulus-absent response. Even though both responses correspond to a no-stimulus report (thus, below the criterion), high-confidence "no" responses were significantly more accurate than low-confidence "no" responses. However, this result is not always obtained. Using several manipulations of stimulus visibility, Kanai and colleagues (2010) asked whether confidence differed for misses and correct rejections. Confidence did not differ with perceptual manipulations of stimulus visibility such as masking. Notably, task accuracy could reach 70 percent before confidence ratings began to distinguish between misses and correct rejections. Interestingly, when perceptual reports were manipulated via attentional load (e.g., cueing or rapid visual presentation), there was a clear relationship between confidence and accuracy for unseen trials. In general, metacognitive sensitivity is worse for stimulus-absent than for stimulus-present responses, even when metacognitive sensitivity is above chance for both (Meuwese et al., 2014; Mazor, Friston, & Fleming, 2020). Lower variance of the stimulus-absent compared to the stimulus-present SDT distribution could also contribute to poorer metacognitive sensitivity for stimulus-absent trials, but it would not predict zero sensitivity (Lau, 2019).

Taking these findings together, it seems that observers at least sometimes have cognitive access to signals from stimuli they report not seeing. These findings can be interpreted in three alternative ways. (1) Conscious perception exists below the criterion for report, violating the predictions of options 2 and 3. (2) Metacognitive processes are sensitive to unconscious information, such that sensory responses below the criterion are unconscious but still capable of influencing confidence ratings. (3) Reports of stimulus presence or absence do not perfectly reflect phenomenology, and observers sometimes report a stimulus as absent when in fact a weak conscious sensation occurred. On this account, metacognitive ability is a product of a weak but nevertheless still conscious experience. The first two alternatives assume that the measured SDT criterion is the phenomenal criterion, whereas the third does not.

To the extent that the sensory stage representation in SDT can be mapped onto activity in early sensory brain areas, a number of neuroimaging and electrophysiology experiments also shed light on the model-phenomenology mapping. Using functional magnetic resonance imaging (fMRI) in a contrast

detection experiment, Ress and Heeger (2003) showed that BOLD responses in the visual cortical area V1 tracked observers' reported perceptions rather than physical stimulus properties. That is, V1 responses were not detectable when the target stimulus was present yet undetected by the observer. In addition, when the observer reported seeing a target that was not present (a false alarm), V1 responded above baseline. Thus, on the basis of these results (and many similar findings from other paradigms; Polonsky et al., 2000; Lee, Blake, & Heeger, 2005; Wunderlich, Schneider, & Kastner, 2005), the sensory response seems to track conscious perception rather well,[5] in favor of option 1.

However, recent work has shown that stimuli that are subjectively invisible and do not afford above-chance discrimination accuracy can nevertheless be decoded from fMRI patterns in V1 (though not in V2 or V3; Haynes & Rees, 2005), suggesting that a mere representation of the stimulus in V1 is not sufficient for reporting awareness. Likewise, recent results from multi-unit recordings in the visual cortex demonstrate that high-contrast targets that are undetected by the animal still elicit large V1 and V4 responses, which are even larger than responses to low-contrast targets that are detected (van Vugt et al., 2018). One interpretation is that there is no threshold level of activity that V1 or V4 could produce that would result in a conscious sensation, suggesting that a mechanism beyond evaluating the response magnitude in visual areas is needed to trigger a report that the stimulus was present. However, the same behavior could be explained by assuming that the animal used a different decision criterion for high- versus low-contrast stimuli, or that the animal's present/absent reports do not perfectly track their phenomenal experience.

SDT has also been used to explain the phenomenon of blindsight, whereby an individual with damage to the primary visual cortex reports blindness in part of the visual field but can still discriminate certain visual stimuli in the "blind" field at above-chance levels (Cowey & Stoerig, 1991). Ko and Lau (2012) argue that in blindsight, the sensory distribution on stimulus-present trials is reduced to just barely above the noise distribution, but the absolute phenomenal criterion remains at higher pre-lesion levels. Thus, blindsight patients almost never report awareness (no threshold crossing) but have some preserved sensitivity (because the post-lesion sensory distribution is slightly above the noise distribution). Under this account, then, the setting of the criterion, and not the sensory response

itself, is what is critical for conscious perception (as in options 2 and 3). This interpretation is concordant with *higher-order* theories of perception in philosophy, which argue that mere sensory representation is not sufficient for consciousness (Brown, Lau, & LeDoux, 2019). Alternatively, it has been argued that blindsight patients are not blind in the sense of lacking phenomenal visual experience, but rather just have severely degraded, albeit still conscious, visual perception (Overgaard et al., 2008), which is below the decisional criterion for report (though see Azzopardi & Cowey, 1997, for an attempt to address the issue of response bias in blindsight). This interpretation is more in line with *first-order* theories in philosophy, which maintain that sensory representation of a particular kind *is* sufficient for conscious perception (Block, 2011), more akin to option 1.

Thus, evidence to date does not definitively rule out any of the possible mappings between SDT components and phenomenology. Later, in section 9.3, we outline future directions that could aid in adjudicating between models and options.

9.2.2 The Drift Diffusion Model

In its simplicity, SDT fails to capture a salient aspect of perceptual decision making, namely that decisions unfold over time. Whereas SDT posits a single internal response on each trial, the DDM assumes that internal responses occur across time and are accumulated until a threshold is reached (Ratcliff et al., 2016). At that point, a decision is made and, importantly, a decision time is specified. That is, DDM has the advantage of providing quantitative predictions of not only the sensitivity and bias of an observer, but also the time it takes to commit to a particular decision. Additionally, recent developments to the DDM allow for estimations of confidence (Kiani & Shadlen, 2009).

Model characterization In the DDM (figure 9.3), an internal response is generated at each moment during stimulus processing (this is sometimes called the momentary evidence). For instance, one can think of the firing rate of a visual neuron in some short time window as the momentary evidence for that neuron's preferred stimulus. A secondary neuron then accumulates that evidence by integrating it across time (perhaps with some leak) until the amount of accumulated evidence surpasses a threshold, whereby a decision is made. The DDM has gained prominence, in part, due to the discovery of neurons in frontal and parietal cortices that ramp up like evidence accumulators when their activity is averaged across trials (Gold &

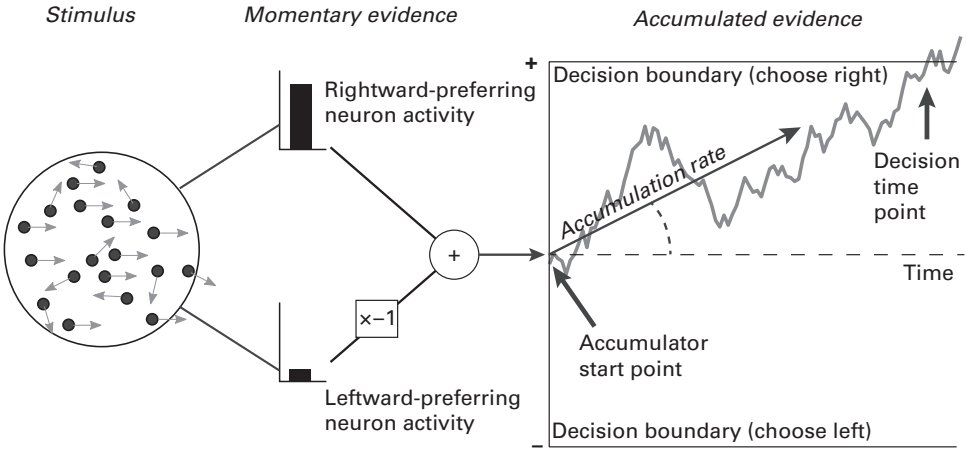


Figure 9.3
Drift diffusion model (DDM) illustrated for a motion-direction discrimination task. Assume a choice between net leftward or rightward motion of some moving dots, where the majority of dots are moving rightward (the stimulus). The sensory representations in DDM begin with activity in feature-tuned detectors (such as those in the middle temporal (MT) area with motion-direction preference). This sensory activity is sustained over time while the stimulus is presented and provides noisy evidence at each time point for how much leftward or rightward motion there is. This is called the momentary evidence. During evidence accumulation, a separate computation integrates the difference of the noisy responses that each feature-tuned detector emits at each moment in time. Thus, the accumulation will build toward a positive or negative boundary. Once this decision variable reaches either boundary, a decision is made. The DDM can thereby explain which choice was made and how long the decision took (reaction time). If the difference in momentary evidence between two detectors is large, this will drive accumulation toward a boundary faster, leading to shorter reaction times. The size of the difference in evidence strength between detectors determines the accumulation rate in the model. The accumulation starting point and placement of the boundaries are also model parameters that can be modified to capture decision biases or speed-accuracy trade-offs, for example.

Shadlen, 2000, 2007; Latimer et al., 2015). A minimal set of parameters for the DDM typically includes the starting point of the accumulator (e.g., zero if there is no bias), the rate of evidence accumulation (sometimes called drift rate), and the position of the decision thresholds (or the bounds). Additionally, when fitting response time data, a non-decision time parameter is often added to account for stimulus-independent sensory- and motor-related latencies in responding.

Sensory stage The sensory representation in the DDM is thought of as a sample of evidence for the presence of a particular stimulus feature at each moment in time. For instance, neurons such as those in the monkey middle temporal (MT) area that fire at a near-constant level while their preferred direction of motion is presented could be considered to generate constant momentary evidence for a particular motion direction. There is a clear relationship between sensory responses (momentary evidence) and accumulated evidence. For example, increasing the contrast of a stimulus or the coherence of motion will drive larger responses in sensory neurons (higher momentary evidence), which will cause an increase in the rate of evidence accumulation, thereby leading to faster threshold crossing and reaction times. This basic intuition explains why easy-to-perceive stimuli are both more accurately and more rapidly discriminated. Typically, in tasks with two alternatives, there is a representation of momentary evidence for one choice and a separate representation of momentary evidence for the other choice, and the difference between these signals feeds into the accumulator (see next section). The sensory signal is often modeled as successive independent draws from a Gaussian distribution with a mean centered on the strength of the momentary evidence (which could be a difference signal) and variance reflecting internal noise.

Decision stage The DDM accumulator can be thought of as a decision variable that, at some point, reaches a threshold, which marks the commitment to a decision. There is much debate, however, as to whether neurons in, for instance, the parietal cortex that have been argued to represent accumulated evidence are more sensory or motor in nature. In early experiments, neurons that behaved as evidence accumulators in the lateral intraparietal (LIP) area of the parietal cortex were chosen specifically because of their motor control functions in driving saccadic behaviors (Gold & Shadlen, 2007). That is, evidence for an accumulating decision variable has mostly been observed in neurons directly involved in the motor output that the animal must produce to indicate a specific percept. This may support an account on which the accumulation process is itself more motor in nature than sensory. However, a recent experiment dissociated more sensory from more motor LIP neurons by placing the saccade response targets in the opposite hemifield to the visual stimuli. Under these conditions, perceptual sensitivity decreased following inactivation of LIP neurons contralateral to the visual stimulus but ipsilateral to the saccade response targets, suggesting that the relevant neurons were more sensory (Zhou & Freedman, 2019).

It remains unclear, though, whether such sensory neurons have similar accumulation dynamics as the motor neurons and to what extent evidence accumulation is obligatory in any perceptual experience.

Although current evidence is not decisive about whether the accumulated evidence should be thought of as sensory, motor, or some intermediate representation, the threshold crossing is considered decisional in nature. Just as for SDT, we can distinguish between *decisional* and phenomenal boundaries. Decisional boundaries (the standard interpretation when fitting DDM models to behavior) can be flexibly adjusted to implement strategic control over one's decisions (e.g., when prioritizing response speed vs. accuracy). We do not expect decision strategy to change perception. So, we can reject a mapping of consciousness onto this decisional boundary crossing. Rather, here, we consider the possibility that there is a less flexible phenomenal boundary, the crossing of which would underlie the formation of a conscious percept.

Where is conscious perception in the DDM? As with SDT, there are several viable options for how DDM components map onto phenomenology. With the addition of a temporal component to DDM, the options make notably distinct predictions about when a stimulus becomes conscious—i.e., how phenomenology evolves over time.

Options for mapping model components to phenomenology

Option 1: Phenomenal perception tracks the sensory representation (the momentary evidence) before it is fed into the accumulation process. This would imply that evidence accumulation and decision commitment are post-perceptual, and perhaps are only engaged when a choice about one's percept needs to be made. According to this option, then, perceptual experience represents stimuli as being a certain way (e.g., seeing dots moving upwards), but this is not enough to trigger an explicit decision that one is seeing the world in that way.

Option 2: Conscious perception tracks the evidence accumulation process. This would imply that conscious experience unfolds as the decision variable diffuses toward a boundary. This would also imply that perceptual experience of stimulus properties becomes stronger over time to the extent that evidence accumulation approaches one or another threshold.

Option 3: Conscious perception occurs once a decision threshold is reached. This would imply that momentary and accumulated evidence are unconscious and that phenomenal experience occurs only once a boundary is reached. It may also imply limited or no conscious access to the state of the

decision variable until the threshold is reached. As with SDT, this option would need to posit a separate and more flexible decisional boundary such that decisions could be made in a perception-independent manner according to other task goals.

Evidence for mappings Despite wide adoption of DDM as a model of perceptual decision making, as well as theoretical discussion of links between consciousness and accumulation processes (Shadlen & Kiani, 2011; Dehaene 2009; Dehaene et al., 2014), there have been surprisingly few empirical attempts to link particular aspects of the DDM to conscious perception of a stimulus. A notable exception is a recent experiment by Kang and colleagues (2017). In this experiment, a motion stimulus with one of two directions and variable coherence (across trials) was presented within an annular boundary surrounding fixation. At the center of fixation, observers also viewed a clock with a rotating hand. Observers were to report the direction of motion followed by the position of the clock hand at the moment they subjectively were aware of having made their choice. By fitting a DDM model using the reported moment of the decision as a substitute for reaction time, the authors could accurately predict observers' choice behavior across coherence levels. The accuracy of these predictions suggests that the DDM model captured the subjective reports of decision timing. Thus, the authors concluded that conscious awareness of having made a choice aligns with the threshold crossing postulated in the DDM, in line with option 3. This is not to say that there was no awareness of the stimulus prior to threshold crossing—observers were clearly seeing something (some moving dots). However, their awareness of the stimulus *as moving in a particular direction* may occur once evidence accumulation terminates at a boundary. Alternatively, this result may only show that observers are aware of when they committed to a decision, rather than indicating the moment when the conscious perception of directional motion occurred.

In humans, a signal thought to be similar to the parietal cortex evidence accumulation process in nonhuman primates is an electrophysiological signature known as the central parietal positivity (CPP; O'Connell, Dockree, & Kelly, 2012). Recently, Tagliabue and colleagues (2019) tracked the CPP during a contrast discrimination task. The contrast level of the stimulus varied across trials, and observers gave both objective responses about which contrast was presented and subjective responses about the visibility of the stimulus. Evidence accumulation, as indexed by the CPP, closely tracked

subjective reports of stimulus awareness rather than the physical stimulus contrast. When trials were selected with equal subjective awareness levels, variation in contrast did not modulate the CPP. Notably, the CPP bears considerable resemblance—and may be identical (Twomey et al., 2015)—to the highly studied P300 electroencephalography component, which has been argued to be a neural marker of conscious perception (Dehaene & Changeux, 2011). The study by Tagliabue and colleagues (2019) is preliminary evidence that conscious perception of a simple contrast-defined stimulus tracks an evidence accumulation signal. More direct evidence that detection depends on accumulation processes consistent with DDM comes from single-neuron recordings from the posterior parietal cortex in a human participant performing a vibrotactile detection task. Neuronal firing rates ramped up for detected stimuli in a way that was related to choice and confidence behavior through a DDM model that assumed awareness occurs at threshold crossing (Pereira et al., 2020), as in option 3.

These findings raise several questions. Is evidence accumulation an index of awareness or only of reported awareness? For instance, using a no-report paradigm, Pereira and colleagues (2020) also found parietal cortex neurons that were sensitive to stimulus intensity, suggesting that part of the activity is conserved when an explicit decision does not need to be made. However, the temporal profile of this activity was notably different from the evidence accumulation signature (see figure 3 of Pereira et al., 2020), perhaps compatible with options 1 or 2. A further question is whether awareness unfolds contemporaneously with evidence accumulation or, as suggested by Kang and colleagues (2017), only once evidence accumulation reaches a threshold. If, as stated in option 3, conscious perception only occurs after boundary crossing, this leads to the counterintuitive prediction that as the decision variable might fluctuate above and below the threshold over time, the associated perceptual experience could appear and disappear in quick succession. Moreover, in many experiments, the momentary evidence signal is not simultaneously measured but rather is expected to be closely related to the accumulated evidence. Therefore, momentary sensory evidence, rather than accumulated evidence, could just as well reflect awareness, or both could be involved.

9.2.3 Probabilistic Population Codes

SDT and DDMs maintain only a point estimate related to the stimulus variable of interest. However, there are reasons to think that sensory representations

contain information about the uncertainty of stimulus variables as well. First, sensory signals are ambiguous. Retinal images, for example, are two-dimensional, whereas the world is three-dimensional. This loss of information means that retinal images are ambiguous with respect to world states. Retinal images therefore contain information that can be used to compute the probabilities of various world states without specifying a unique state. Second, sensory signals are noisy, or variable. Given repeated presentations of the same stimulus, the brain responds differently each time. The ambiguity and noise in sensory signals have motivated the idea that perception is an inference process (Helmholtz, 1856). That is, the perceptual system is trying to infer what things in the world most likely caused the image on the retina, given the uncertainty of the sensory signals (Knill & Richards, 1996; Kersten, Mamassian, & Yuille, 2004).

It would be advantageous, then, for the system to estimate and use its own uncertainty. We have empirical evidence that it does so. The incorporation of uncertainty into perceptual decisions has been found in many behavioral studies (Trommershäuser, Kording, & Landy, 2011; Ma & Jazayeri, 2014) and shows that the representations used for perceptual decision making are richer than point estimates. To illustrate how this might play out in an everyday activity, consider the following: when driving at night, you might be less certain about the distance between your car and the car in front of you than you would be during the day, and you might therefore keep further back as a result. A point estimate of the distance could be the same at night and during the day, but the greater uncertainty at night would lead to more cautious behavior. Behavioral studies have found both optimal and suboptimal perceptual decision making (Rahnev & Denison, 2018). However, probabilistic information can be incorporated into perceptual decisions even if the outcome is not strictly optimal (Ma, 2012).

PPC and sampling (section 9.2.4) are two widely used approaches for modeling sensory uncertainty. Both are formulated as neural models and describe how neurons could represent probability distributions over stimulus features.

Model characterization In PPC (Jazayeri & Movshon, 2006; Ma et al., 2006), a population of neurons encodes a probability distribution over some sensory feature (Földiák, 1993; Sanger, 1996; Zemel, Dayan, & Pouget, 1998). For example, a population of orientation-tuned neurons would encode a probability

distribution over stimulus orientation. PPC is a static model—it uses a population of neurons to encode a probability distribution at a single moment in time (or at steady state), and it has no time-varying internal processes.

Sensory stage To describe the sensory stage of the PPC model, let's consider a population of orientation-tuned neurons. Each neuron has a tuning function that describes the mean response of the neuron when a stimulus of a particular orientation is presented (figure 9.4). However, the neuron's response is not perfectly predictable because of response variability. This variability reflects uncertainty in the stimulus representation. Response variability is also called noise. Biological neurons are often characterized as having Poisson noise—a particular noise distribution in which the response variance scales with the response mean. PPC assumes Poisson (or Poisson-like) noise, which turns out to have important theoretical consequences for population codes (Ma et al., 2006). The responses of a population of orientation-tuned neurons with Poisson noise can be mathematically combined using a Bayesian decoder to compute a probability distribution over the stimulus orientation (figure 9.4). Specifically, a log transform of the encoded probability distribution can be read out from the population using a weighted, linear combination of each neuron's response. To combine the responses in a Bayesian fashion, two pieces of information are required: (1) the orientation tuning functions of each neuron and (2) the noise distribution. Therefore, in PPC models, the response of a single neuron does not represent a probability. Rather, neural populations can encode probability distributions over stimulus features, where specific mathematical operations are required to calculate probabilities from the population activity.

Decision stage Does the brain actually do the required calculations to compute probability distributions over sensory features? If so, this would be a job for the decision stage. Although PPC usually refers to the sensory encoding stage, there have been various ideas about how information could be read out from the sensory population code at a decision stage. The way such a readout is expected to work in the brain is that sensory neurons input to decision-related neurons, and the decision neurons integrate the sensory inputs. The specific way that the decision stage would integrate sensory inputs could vary, and here we consider different options.

First, and as introduced above, a log transform of the encoded probability distribution can be read out from the PPC using a linear operation. The

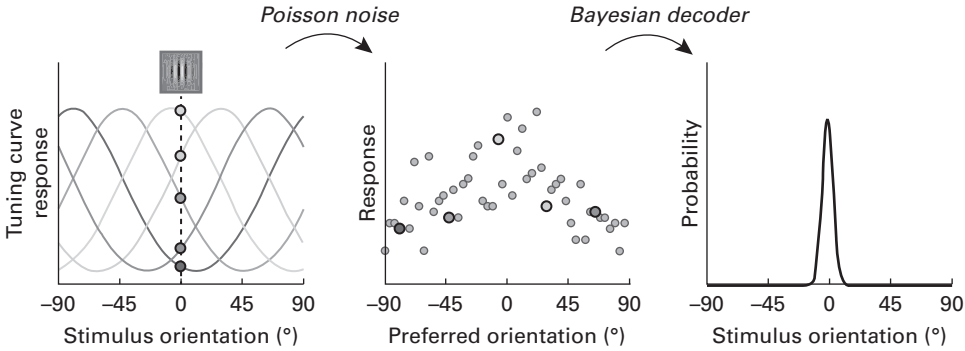


Figure 9.4

Probabilistic population codes. Left: A population of orientation-tuned neurons has tuning curves that tile orientation. Example tuning curves from the population are shown (curves), along with their mean responses to a stimulus with 0° (i.e., vertical) orientation (points). Middle: On a single trial, the response of each neuron (gray dots) is determined by its tuning curve and Poisson noise; the preferred orientation of each neuron is the orientation at which its tuning curve peaks. Right: A probability distribution can be read out from the single-trial population response using a Bayesian decoder, which combines the information from all neurons to calculate the likelihood of the population response given a stimulus of each orientation. See color plate 2.

linearity makes it plausible that a biological neural network could perform this computation, and specific proposals have been made for how it might do so (Jazayeri & Movshon, 2006; Ma et al., 2006; Beck, Ma, et al., 2008; Beck, Pouget, & Heller, 2012; Ma & Jazayeri, 2014). In this operation, each neuron's response is weighted by the log of its tuning function value at that stimulus value, and these weighted responses are then summed across the population to compute the log probability. This operation recovers the full probability distribution. Such information would enable optimal decisions in, for example, a two-choice discrimination task by simply comparing the probabilities associated with the two stimulus options. Confidence has been proposed to be based on the probability that the choice is correct (Meyniel, Sigman, & Mainen, 2015; Pouget, Drugowitsch, & Kepecs, 2016), though this idea has not been fully supported empirically (Adler & Ma, 2018). It could also be possible to use the PPC sensory representation to compute probability-dependent expected values or make probabilistic categorical decisions without reconstructing a full probability distribution over stimulus feature values.

Whether and under what circumstances downstream regions would read out a full probability distribution from sensory regions is unknown.

Second, key properties of the probability distribution encoded by the PPC can be extracted from the population activity without reading out the full distribution. For Gaussian distributions, the mean (and max) of the distribution is approximated by the stimulus value at which the population activity peaks, and the variance is inversely proportional to the height of this peak (Ma et al., 2006; figure 9.4). So, a simple downstream readout mechanism could estimate the stimulus value from the peak position and uncertainty from the peak amplitude or from the sum of the amplitudes across the population (Meyniel et al., 2015). Under such a readout, in a two-choice discrimination task, the stimulus option closer to the estimated value would be chosen, and confidence would be based on the uncertainty estimate.

In summary, PPC models show how probabilistic information could be encoded in sensory populations and read out from decision populations. Ongoing empirical work is actively testing whether real neural populations behave like the theorized PPC (Beck, Ma, et al., 2008; Fetsch et al., 2012; Bays, 2014; van Bergen et al., 2015; Hou et al., 2019; Walker et al., 2020) and whether humans have behavioral access to full probability distributions (Rahnev, 2017; Yeon & Rahnev, 2020).

Where is conscious perception in PPC? In PPC models, the sensory representation is more complex than the sensory stage in SDT and DDM models. As a result, there are more possibilities for how the sensory representation itself could be realized in phenomenology (option 1). Alternatively, we can consider mappings between decision stage readouts and phenomenology (option 2).

Options for mapping model components to phenomenology

Option 1: Phenomenal experience maps to the sensory-stage population code. The content of conscious experience is specifically related to neural activity in the neural population that encodes the relevant probability distribution. Here, there are several alternatives for how phenomenal experience could relate to the population activity.

Option 1a: No transformation of the population code. We experience the population activity itself. This might look like an imprecise percept with graded strengths across a continuum of feature values, but those strengths would not be probabilities. If this is the case, then probabilistic information

is not directly available in phenomenal experience because a further transformation of the population activity is required to calculate the probability distribution.

Option 1b: Phenomenology reads out summary statistics of the population code. For example, we could experience a point estimate (such as the mean value represented by the population activity), a point estimate with some strength (such as the mean value plus the gain, or height of the peak), or a point estimate with some uncertainty estimate (Rahnev, 2017). If we experience just the mean, say, our conscious experience of some stimulus feature would consist of that single exact feature value.

Option 1c: Phenomenology consists of a probability distribution derived from the population code. We can divide this into two further sub-options. The first of these is that we experience the (log?) probability distribution. In this case, we could experience the distribution as probabilistic in the sense that our perceptual experience would simultaneously represent multiple alternatives with different likelihoods. This is the option that seems most compatible with the philosopher John Morrison's description of his "perceptual confidence" view (Morrison, 2016). In the second sub-option, alternatively, we could see the distribution but not see it as probabilistic (i.e., we could see all the possible values as simultaneously present with different strengths but not as alternative options about the state of the world). For example, if our vision is blurry, a dot may look smeared out in space, but we don't interpret the edge of the smear as a low probability dot location (Denison, 2017).

These options for translating the PPC into phenomenal experience are intended to span the range from minimal transformation to full probability distribution, but they are not exhaustive; there are many possible ways to transform PPC activity. It is also important to emphasize that in options 1b and 1c, summary statistics and probabilities are not explicitly represented in the sensory stage population itself.[6] Rather, the calculation of these quantities occurs (somehow) in the translation to phenomenal experience.

Option 2: Although a separate decision stage is not part of the PPC model, we can think of a second option wherein the information in the sensory-stage population code is read out by a decision stage, and only this explicit representation becomes conscious. Similar alternatives as described for option 1 would apply, but option 2 requires a decision stage in the brain to read out

(i.e., explicitly compute) summary statistics, probability distributions, and so on from the sensory stage activity.

Evidence for mappings PPC models have rarely been related to conscious phenomenology. One exception is a study reporting that perceptual rivalry—in which perception of an ambiguous stimulus alternates between two perceptual interpretations—is consistent with Bayesian sampling from PPCs (Moreno-Bote, Knill, & Pouget, 2011; see also *Evidence for Mappings* in section 9.2.4). Lehky and Sejnowski (1999) have discussed phenomenal experience in relation to population coding models more generally, making the point that our subjective experience is not determined by physical variables as a physicist would describe them but rather by those physical variables transformed according to the ways in which our perceptual apparatus interacts with them. For example, our perception of color requires a transformation from physical variables involving wavelengths and reflectances into the psychological variable of hue.

Though clear links have not been made to phenomenal experience per se, testing whether PPC models can link behavioral performance to neural activity is an active area of research. One study decoded orientation uncertainty from fMRI responses in human visual cortex, and showed that this uncertainty correlated with both the variability and biases in perceptual decision behavior (van Bergen et al., 2015). Decoded likelihood functions from neural populations in the primate visual cortex likewise could predict trial-by-trial fluctuations in decision behavior (Walker et al., 2020). Therefore, the uncertainty information measured in PPC appears to be used in behavior. This fact means that the represented information is more than a single point estimate (see option 1b). Note, however, it does not mean that the full probability distributions are used, as required by option 1c (Rahnev, 2017). And it remains an open question whether the information used in behavior in these cases is also reflected in perceptual experience.

Some philosophers have argued that perceptual experience is probabilistic (Morrison, 2016; Munton, 2016; Morrison, 2017), in the sense that our conscious perceptual experience of a feature value includes the probability of that feature value (option 1c, sub-option 1). However, others have argued that experience is not probabilistic (all other options; Denison, 2017; Block, 2018; Beck, 2019; Cheng, 2018; Pohl, S., Perceptual representations are not

probabilistic, unpublished). The extent to which probabilistic information is explicitly represented in the brain is relevant for this debate. Recall that at the sensory stage of PPC models, probability is encoded only implicitly; it requires a further step to be explicitly computed. That is, the explicit number corresponding to the probability of a given stimulus feature value can only be calculated by combining the neural responses in a specific way. There is no guarantee that such a calculation ever occurs; and if it does, it's not known whether it would have phenomenal content (Lehky & Sejnowski, 1999). A philosophical view that argues for probabilistic perception based on PPC models would need to specify how the computation of probability occurs—either somehow in the direct transformation of sensory representations to phenomenology (option 1) or explicitly in a separate decision stage (option 2).

9.2.4 Sampling

Whereas PPC models represent probability distributions through the activity of a large number of neurons at a single time, sampling models represent probability distributions through the activity of single neurons across time (Fiser et al., 2010). Therefore, unlike PPC models, sampling models are inherently dynamic.

Model characterization Sampling models are strongly motivated by the idea of perception as probabilistic inference. In these models, neuronal response variability is a feature rather than a bug—it's a way to represent uncertainty over ambiguous world states. Sampling models propose that single neurons represent posterior probabilities of states of the world. They do so through a sequence of samples from the posterior probability distribution they represent (Hoyer & Hyvärinen, 2003).

Sensory stage Each neuron corresponds to some stimulus property. A neuron's response represents a probability distribution over the possible values of that property. At an early sensory stage, these properties could be, for example, the intensity of an oriented patch[7] present in a stimulus at some location (Olshausen & Field, 1996; Hoyer & Hyvärinen, 2003; Haefner, Berkes, & Fiser, 2016; Orbán et al., 2016), and a given neuron's response might represent the intensity of a vertical orientation. Specifically, it would represent the probability of the vertical intensity of the stimulus given the image. The response of a neuron at a given time (e.g., its instantaneous firing rate)

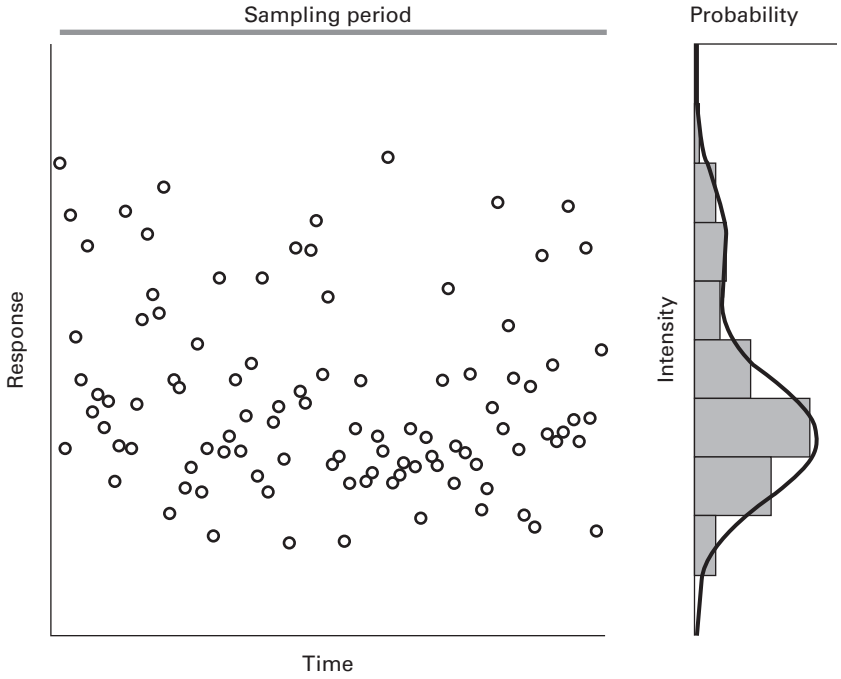


Figure 9.5
Sampling models. Left: The response (e.g., instantaneous firing rate) of a single neuron is plotted at successive time points. Right: The responses of this neuron are drawn from a probability distribution (black curve), such that the responses aggregated across the sampling period (gray histogram) approximate the probability distribution. If this neuron represents the intensity of a vertical stimulus, this series of responses indicates that it is most likely that the vertical intensity is low, but there is some probability that the intensity is higher.

is modeled as a sample from the posterior distribution over possible stimulus values. In our example, if there is high probability that the vertical intensity is low, the firing rate of the neuron will mostly be low. Samples collected over time will build up the probability distribution (figure 9.5). Given enough time, the probability distribution can be approximated with arbitrary precision. Recent theoretical work explores how sampling models could be implemented in biologically realistic neural networks (Buesing et al., 2011; Huang & Rao, 2014; Savin & Denève, 2014; Petrovici et al., 2016).

Decision stage In sampling models, there is not a sharp distinction between sensory and decision stages because all neurons represent posterior probabilities

over world state variables. Decision making in sampling models therefore only requires reading the activity of neurons that explicitly represent the relevant variable for the task.

For example, in an orientation discrimination task, we imagine that there are neurons downstream of early sensory areas that correspond to decision variables: the presence of a counterclockwise stimulus and the presence of a clockwise stimulus (Haefner et al., 2016). These neurons would learn the posterior distributions over their decision variables from task training (Fiser et al., 2010). Then, the final decision computation would be simply to compare the counterclockwise response to the clockwise response over some length of time. This accumulation of evidence across time could resemble a drift diffusion process (Yang & Shadlen, 2007; Haefner et al., 2016).

Where is conscious perception in sampling? For sampling models, as for PPC models, the complexity of the proposed representation makes multiple mappings possible. However, in sampling models, the complexity arises from a fundamentally different type of neural code, with its own mapping options.

Options for mapping model components to phenomenology In sampling models, the responses of all neurons are samples from posterior distributions. So, the models imply that the content of conscious perception is based on these samples. There are several different options for how neural responses in sampling models could map to conscious perception.

Option 1: Our phenomenal experience is based on one sample: we experience the current sample. This option says that our conscious perception has no probabilistic content and only reflects the brain's current guess about the state of the world. This option implies that our perceptual experience changes with every new sample. Because neural activity is constantly changing and different parts of the brain receive information about the same stimulus at different times, it also requires a mechanism for samples across the brain to be synchronized in phenomenal experience: experience the current guess from all neurons in the brain *now*.

Option 2: Our phenomenal experience is based on multiple samples, which could be interpreted in one of two ways:

Option 2a: We experience a summary statistic based on samples across some time interval. For example, we may experience the mean value of the samples. In this case, our conscious perception would still be a point

estimate. However, this option could also allow perception of uncertainty by extracting the variability across samples over the time interval.

Option 2b: We experience an approximated probability distribution based on samples across some time interval. We could either experience this distribution as probabilities, in which case, our phenomenal experience would be probabilistic, or we could see it as non-probabilistic—that is, we could simultaneously experience the different values of the distribution without experiencing these values as competing options.

Evidence for mappings There have been a few attempts to relate sampling models to conscious perception. Perhaps the most explored idea is that sampling models are well suited to predict the alternating perceptual interpretations observers see when viewing bistable images (Sundareswara & Schrater, 2008; Gershman & Tenenbaum, 2009; Moreno-Bote et al., 2011; Gershman, Vul, & Tenenbaum, 2012). In this case, the probability distribution over the world state is bimodal, and sampling models would sequentially sample from each of the modes of the distribution. Moreno-Bote and colleagues (2011) used sampling from PPCs to model perceptual alternations, combining the two probabilistic approaches. Recent theoretical work has described a unification of sampling and PPC models, showing that these probabilistic approaches are not mutually exclusive (Shivkumar et al., 2018). It has also been argued that perceptual reports during spatial and temporal selection tasks reflect probabilistic samples (Vul, Hanus, & Kanwisher, 2009) and that cognition more broadly reflects a Bayesian sampling process (Sanborn & Chater, 2016).

Testing perceptual sampling models in neural systems is in the early stages. Spontaneous activity in the visual cortex resembles evoked activity to natural images, which has been argued to be consistent with a prediction from sampling models. In the absence of stimulus-driven activity, neural responses will reflect the prior (in this case, the statistics of natural images; Berkes et al., 2011). Various statistical properties of responses in the visual cortex during visual stimulation and visual tasks are also consistent with sampling models (Haefner et al., 2016; Orbán et al., 2016; Bondy et al., 2018; Banyai et al., 2019).

When evaluating the relation between sampling models and phenomenology, it is important to keep in mind that even if phenomenal experience arises from a single sample or is a point-estimate summary statistic, perceptual decision behavior could still reflect uncertainty because (1) the current sample reflects the posterior probability distribution, which already incorporates the relative uncertainties of priors and likelihoods, and (2) decision readout mechanisms could still be based on multiple samples across time.

The hypothesis that phenomenal experience is based on multiple samples—a possibility offered only by the sampling model—raises several questions: Under this hypothesis, how many samples are required to experience something? Is our phenomenal experience always updated after a fixed number of samples? If not, how does the system know when to start a new round of sampling? Are successive percepts based on independent sets of samples, or a moving window of samples? Future empirical work could pursue these questions and perhaps, in the process, determine how well sampling models can explain and predict phenomenology.

9.3 Conclusions and Future Directions

9.3.1 Constrained, but not Unique, Mappings between Model Components and Phenomenal Experience

We considered four computational models—SDT, DDM, PPC, and sampling—which, taken together, have been broadly applied to perceptual decision making and perceptual coding in the brain. For each model, we find no unique mapping between model components and phenomenal experience. Rather, we find that each model is consistent with multiple possible mappings. Even SDT, the simplest model, affords different ways of relating internal responses, criteria, and their combination to phenomenal experience.

Importantly, the models cannot support any arbitrary mapping to phenomenal experience. Instead, each model offers a constrained set of possible mappings. The models we discussed here differ in three major ways in terms of the mappings they can support. First, the models differ in their uses of thresholds. SDT and DDM models use thresholds to define decision behavior, whereas PPC and sampling do not include thresholds and are more focused on the nature of the sensory representation. (This difference is perhaps related to the fact that SDT and DDM are specifically designed to model task-based decision behavior, whereas PPC and sampling are not.) In models with thresholds, different mappings predict that consciousness either is or is not tied to a threshold crossing. This distinction is relevant to ongoing debates about whether consciousness is all or nothing and whether an "ignition" process underlies conscious awareness of specific content (Dehaene, Sergent, & Changeux, 2003; Sergent & Dehaene, 2004; Fisch et al., 2009; Vul et al., 2009; Moutard, Dehaene, & Malach, 2015; Noy et al., 2015; van Vugt et al., 2018). Second, the models differ in how they represent uncertainty. PPC models could support a phenomenal experience of uncertainty

instantaneously, whereas sampling and DDM models could only do so over time, and SDT models (with their single point estimate) could not do so at all. (NB: Whether we have a perceptual experience of uncertainty in addition to our cognitive sense of it is up for debate; see *Evidence for Mappings* in section 9.2.3.) Third, the models can support different kinds of predictions about the timing of phenomenal experience. In DDM, different mappings make distinct predictions about when consciousness occurs (before or after boundary crossing) and how gradual it is (tracking continuous momentary evidence, continuous evidence accumulation, or all or nothing once a boundary is reached). In sampling, different mappings can support phenomenal experience either for each sample or only after a period of sampling. The static models, SDT and PPC, don't afford predictions about timing.

The lack of unique mappings between model components and phenomenal experience is in some sense not surprising; these models were developed to predict perceptual decision behavior and brain activity rather than phenomenal experience. This gap raises two questions: What approaches can we take, in science and in philosophy, to develop tighter mappings between vision models and visual experience? And given the private nature of subjective experience, how far can we hope to get with such a project?

9.3.2 Directions for Science

We see considerable opportunity for advancing our understanding of the processes that underlie phenomenal experience using computational models. We can do this both by collecting more and different kinds of subjective reports in experiments and by developing our models to expand the kinds of reports they can predict. These two research directions work in tandem. Our goal should be vision models that predict a variety of objective and subjective perceptual reports.

There are several well-established psychophysical protocols for collecting subjective reports. Subjective reports can be classified as either reports about the observer's phenomenal experience of the stimulus or reports about the observer's objective decision. Subjective reports about phenomenal experience include visibility ratings, appearance judgments, and appearance matching. For example, an observer may be asked to report whether a test stimulus appears higher contrast than a stimulus with a standard contrast (Carrasco, Ling, & Read, 2004). This protocol allows estimation of the point of subjective equality: the contrast required for the observer to report that the test contrast is higher than the standard 50 percent of

the time. Methods for measuring the perceived magnitude (e.g., brightness, loudness) of suprathreshold stimuli are called scaling methods.

Subjective reports about the observer's decision include confidence ratings, confidence comparisons (Barthelmé & Mamassian, 2009), and post-decision wagering (Persaud, McLeod, & Cowey, 2007). For example, an observer may be asked to perform an objective perceptual task and then report their confidence as the likelihood that their objective choice was correct. The topic of perceptual confidence has seen a surge of research interest in recent years (Mamassian, 2016; Pouget et al., 2016), including work modeling the computations underlying confidence reports (Kiani & Shadlen, 2009; Fetsch, Kiani, & Shadlen, 2014; Meyniel et al., 2015; Adler & Ma, 2018; Denison et al., 2018; Ott, Masset, & Kepecs, 2019). In contrast, appearance reports have received much less attention and little modeling. This is true, despite their critical role in the history of vision science. Perceptual matching tasks were used, for example, to establish the trichromatic theory of color vision and to demonstrate that perceived contrast is approximately constant across the visual field (Georgeson & Sullivan, 1975). Arguably, appearance reports allow more direct inferences about the content of phenomenal experience than metacognitive reports about the perceptual decision. We see particular potential in extending our current models to capture appearance comparison and similarity judgments. One promising current direction is maximum likelihood difference scaling—an approach for collecting and modeling suprathreshold appearance judgments (Maloney & Knoblauch, 2020).

9.3.3 Directions for Philosophy

We consider two directions for philosophy in this topic area. First, directly relevant to scientific practice, philosophers could consider the questions: (1) What should we try to explain and predict about phenomenal experience? (2) What can we hope to explain and predict about phenomenal experience? To make scientific progress on the nature of phenomenal experience, we must use the data of behavioral reports and neural measurements. In this way, inferring values on dimensions of phenomenal experience is a theoretical enterprise similar to others in science, such as inferring the masses of stars or the structures of atoms. Can we approach modeling phenomenal experience in the same way that we approach modeling these other entities? Or are there special considerations for modeling phenomenal experience? Are there any hard limits on what we can learn about phenomenal experience from behavioral and neural data, and if so,

what are they? What is the role of neural data, in particular, in learning about phenomenal experience, and how can we best combine behavioral and neural data to make inferences about experience? The models we have considered are usually applied to perceptual decisions about a single visual feature. Is there a way to bridge to multiple simultaneous features or even visual phenomenology as a whole?

Second, as mentioned, our current models, if taken literally as models of phenomenal experience, would make very different predictions for what information about perceptual uncertainty could be available in phenomenal experience. The question of whether and how we may perceptually experience uncertainty has recently begun to attract interest in philosophy (see *Evidence for Mappings* in section 9.2.3). It is relevant for understanding not only the nature of experience (philosophy of mind; Morrison, 2016) but also whether perception must be probabilistic in order to justify probabilistic beliefs about the world (epistemology; Munton, 2016; Nanay, 2020). Currently, it is not even clear what the various options are for experiencing perceptual uncertainty in principle. We see this as an open direction in philosophy. The more such work in philosophy is able to relate to scientific models and data from perception science, the more productive we believe this line of work will be for both disciplines.

Acknowledgments

We would like to thank the following individuals who provided valuable feedback on earlier versions of this chapter: Ralf Haefner, Luke Huszar, Richard Lange, Brian Maniscalco, Matthias Michel, Jorge Morales, Dobromir Rahnev, Adriana Renero, Bas Rokers, Susanna Siegel, and Karen Tian. Thanks also to the attendees of the 2019 Summer Seminar in Neuroscience and Philosophy (SSNAP) at Duke University and the Philosophy of Mind Reading Group at New York University for helpful discussions.

Notes

1. Although the machinery of these models is dynamic, they are typically used to represent static stimuli or stable stimulus properties.

2. Although a single number could be used to represent a probability computed at a previous stage, here we use "probabilistic" to refer to the representational stage at which the probability is computed.

3. Philosophers distinguish between the core and total neural realizers of a conscious state. A total neural realizer is sufficient for a conscious state, but a core realizer is sufficient only given background conditions. The core realizer of the experience of red differs from the core realizer of the experience of green, but the background conditions that have to be added to either of these core realizers to obtain a sufficient condition of a conscious state may be the same (Shoemaker, 2007). Neural activity in the middle temporal (MT) cortical area correlates with perception of motion, but no one should think that if MT were removed from a brain and kept alive that activations in it would constitute motion experiences. Activity in MT is a core realizer of the experience of motion, but background conditions need to be added to it to obtain a total realizer.

A further important distinction is between background conditions that are constitutive of a total realizer of consciousness and background conditions that are contingently related to consciousness. Blood flow in the brain is a contingent background condition—you cannot have sustained neural activity without it—but there might be some other way of providing oxygen to neurons. Corticothalamic oscillations may be a constitutive background condition.

4. This prediction also assumes that stimulus-absent and stimulus-present reports faithfully reflect the observer's conscious experience.

5. One caveat of interpreting fMRI BOLD activity in a given brain region as a sensory response is that feedback from higher-order regions may contribute to it.

6. Different authors may use "explicit" differently. Pouget and colleagues (2016) call a representation explicit as long as it is linearly decodable. Here, we use "explicit" to mean that a neuron's response is proportional to the value of interest. One could also argue that representing the parameters of a probability distribution is sufficient for representing the full distribution, but here we intend that a full distribution is "explicitly" represented only when the probabilities associated with any given feature value map to some neuron's firing rate.

7. The patch could reflect multiple stimulus properties, including location, orientation, spatial frequency, color, disparity, and so on. In our example, we focus on orientation for simplicity.

References

Adler, W. T., & Ma, W. J. (2018). Comparing Bayesian and non-Bayesian accounts of human confidence reports. *PLoS Computational Biology, 14*, e1006572.

Azzopardi, P., & Cowey, A. (1997). Is blindsight like normal, near-threshold vision? *Proceedings of the National Academy of Sciences of the United States of America, 94*, 14190–14194.

Banyai, M., Lazar, A., Klein, L., Klon-Lipok, J., Stippinger, M., Singer, W., & Orban, G. (2019). Stimulus complexity shapes response correlations in primary visual

cortex. *Proceedings of the National Academy of Sciences of the United States of America, 116*, 2723–2732.

Barthelmé, S., & Mamassian, P. (2009). Evaluation of objective uncertainty in the visual system. *PLoS Computational Biology, 5*, e1000504.

Bays, P. M. (2014). Noise in neural populations accounts for errors in working memory. *Journal of Neuroscience, 34*, 3632–3645.

Beck, J. (2019). On perceptual confidence and "completely trusting your experience." *Analytic Philosophy, 61*, 174–188.

Beck, J. M., Ma, W. J., Kiani, R., Hanks, T., Churchland, A. K., Roitman, J., . . . Pouget, A. (2008). Probabilistic population codes for Bayesian decision making. *Neuron, 60*, 1142–1152.

Beck, J. M., Pouget, A., & Heller, K. A. (2012). Complex inference in neural circuits with probabilistic population codes and topic models. *Advances in Neural Information Processing Systems, 25*, 3059–3067.

Berkes, P., Orban, G., Lengyel, M., & Fiser, J. (2011). Spontaneous cortical activity reveals hallmarks of an optimal internal model of the environment. *Science, 331*, 83–87.

Block, N. (2011). Perceptual consciousness overflows cognitive access. *Trends in Cognitive Sciences, 15*, 567–575.

Block, N. (2018). If perception is probabilistic, why does it not seem probabilistic? *Philosophical Transactions of the Royal Society B: Biological Sciences, 373*, 20170341.

Bondy, A.G., Haefner, R.M., Cumming, B.G. (2018). Feedback determines the structure of correlated variability in primary visual cortex. *Nature Neuroscience, 21*, 598–606.

Brown, R., Lau, H., & LeDoux, J. E. (2019). Understanding the higher-order approach to consciousness. *Trends in Cognitive Sciences, 23*, 754–768.

Buesing, L., Bill, J., Nessler, B., & Maass, W. (2011). Neural dynamics as sampling: A model for stochastic computation in recurrent networks of spiking neurons. *PLoS Computational Biology, 7*, e1002211.

Carrasco, M., Ling, S., & Read, S. (2004). Attention alters appearance. *Nature Neuroscience, 7*, 308–313.

Chakravartty, A. (2017). Scientific realism. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Summer 2017 ed.). Retrieved from https://plato.stanford.edu/archives/sum2017/entries/scientific-realism/.

Cheng, T. (2018). Post-perceptual confidence and supervaluative matching profile. *Inquiry*, 1–29.

Cowey, A., & Stoerig, P. (1991). The neurobiology of blindsight. *Trends in Neurosciences, 14*, 140–145.

Dehaene, S. (2009). Conscious and nonconscious processes: Distinct forms of evidence accumulation? *Séminaire Poincareé, XII*, 89–114.

Dehaene, S., & Changeux, J.-P. (2011). Experimental and theoretical approaches to conscious processing. *Neuron, 70*, 200–227.

Dehaene, S., Charles, L., King, J.-R., & Marti, S. (2014). Toward a computational theory of conscious processing. *Current Opinion in Neurobiology, 25*, 76–84.

Dehaene, S., Sergent, C., & Changeux, J.-P. (2003). A neuronal network model linking subjective reports and objective physiological data during conscious perception. *Proceedings of the National Academy of Sciences of the United States of America, 100*, 8520–8525.

Denison, R. N. (2017). Precision, not confidence, describes the uncertainty of perceptual experience: Comment on John Morrison's "Perceptual Confidence." *Analytic Philosophy, 58*, 58–70.

Denison, R. N., Adler, W. T., Carrasco, M., & Ma, W. J. (2018). Humans incorporate attention-dependent uncertainty into perceptual decisions and confidence. *Proceedings of the National Academy of Sciences of the United States of America, 115*, 11090–11095.

Fetsch, C. R., Kiani, R., & Shadlen, M. N. (2014). Predicting the accuracy of a decision: A neural mechanism of confidence. *Cold Spring Harbor Symposia on Quantitative Biology, 79*, 185–197.

Fetsch, C. R., Pouget, A., Deangelis, G. C., & Angelaki, D. E. (2012). Neural correlates of reliability-based cue weighting during multisensory integration. *Nature Neuroscience, 15*, 146–154.

Fisch, L., Privman, E., Ramot, M., Harel, M., Nir, Y., Kipervasser, S., . . . Malach, R. (2009). Neural "ignition": Enhanced activation linked to perceptual awareness in human ventral stream visual cortex. *Neuron, 64*, 562–574.

Fiser, J., Berkes, P., Orbán, G., & Lengyel, M. (2010). Statistically optimal perception and learning: From behavior to neural representations. *Trends in Cognitive Sciences, 14*, 119–130.

Földiák, P. (1993). The "ideal homunculus": Statistical inference from neural population responses. In F. H. Eeckman & J. M. Bower (Eds.), *Computation and neural systems* (pp. 55–60). New York: Springer.

Galvin, S. J., Podd, J. V., Drga, V., & Whitmore, J. (2003). Type 2 tasks in the theory of signal detectability: Discrimination between correct and incorrect decisions. *Psychonomic Bulletin and Review, 10*, 843–876.

Geisler, W. S. (2011). Contributions of ideal observer theory to vision research. *Vision Research, 51*, 771–781.

Georgeson, M. A., & Sullivan, G. D. (1975). Contrast constancy: Deblurring in human vision by spatial frequency channels. *Journal of Physiology, 252*, 627–656.

Gershman, S. J., & Tenenbaum, J. (2009). Perceptual multistability as Markov chain Monte Carlo inference. *Advances in Neural Information Processing Systems, 22*, 611–619.

Gershman, S. J., Vul, E., & Tenenbaum, J. B. (2012). Multistability and perceptual inference. *Neural Computation, 24*, 1–24.

Gold, J. I., & Shadlen, M. N. (2000). Representation of a perceptual decision in developing oculomotor commands. *Nature, 404*, 390–394.

Gold, J. I., & Shadlen, M. N. (2007). The neural basis of decision making. *Annual Review of Neuroscience, 30*, 535–574.

Haefner, R. M., Berkes, P., & Fiser, J. (2016). Perceptual decision-making as probabilistic inference by neural sampling. *Neuron, 90*, 649–660.

Haynes, J.-D., & Rees, G. (2005). Predicting the orientation of invisible stimuli from activity in human primary visual cortex. *Nature Neuroscience, 8*, 686–691.

Helmholtz, H. (1856). *Treatise on physiological optics*. London: Thoemmes Continuum.

Hou, H., Zheng, Q., Zhao, Y., Pouget, A., & Gu, Y. (2019). Neural correlates of optimal multisensory decision making under time-varying reliabilities with an invariant linear probabilistic population code. *Neuron, 104*, 1010–1021.e10.

Hoyer, P. O., & Hyvärinen, A. (2003). Interpreting neural response variability as Monte Carlo sampling of the posterior. *Advances in Neural Information Processing Systems, 15*, 293–300.

Huang, Y., & Rao, R. P. (2014). Neurons as Monte Carlo samplers: Bayesian inference and learning in spiking networks. *Advances in Neural Information Processing Systems, 27*, 1943–1951.

Iemi, L., Chaumon, M., Crouzet, S. M., & Busch, N. A. (2017). Spontaneous neural oscillations bias perception by modulating baseline excitability. *Journal of Neuroscience, 37*, 807–819.

Jazayeri, M., & Movshon, J. A. (2006). Optimal representation of sensory information by neural populations. *Nature Neuroscience, 9*, 690–696.

Kang, Y. H. R., Petzschner, F. H., Wolpert, D. M., & Shadlen, M. N. (2017). Piercing of consciousness as a threshold-crossing operation. *Current Biology, 27*, 2285–2295.

Kanai, R., Walsh, V., & Tseng, C.-H. (2010). Subjective discriminability of invisibility: A framework for distinguishing perceptual and attentional failures of awareness. *Consciousness and Cognition, 19*, 1045–1057.

Kersten, D., Mamassian, P., & Yuille, A. (2004). Object perception as Bayesian inference. *Annual Reviews in Psychology, 55*, 271–304.

Kiani, R., & Shadlen, M. N. (2009). Representation of confidence associated with a decision by neurons in the parietal cortex. *Science, 324*, 759–764.

Knill, D. C., & Richards, W. (Eds.). (1996). *Perception as Bayesian inference*. New York: Cambridge University Press.

Ko, Y., & Lau, H. (2012). A detection theoretic explanation of blindsight suggests a link between conscious perception and metacognition. *Philosophical Transactions of the Royal Society B: Biological Sciences, 367*, 1401–1411.

Koeing, D., & Hofer, H. (2011). The absolute threshold of cone vision. *Journal of Vision, 11*, 21.

Kupers, E. R., Carrasco, M., & Winawer, J. (2019). Modeling visual performance differences "around" the visual field: A computational observer approach. *PLoS Computational Biology, 15*, e1007063.

Latimer, K. W., Yates, J. L., Meister, M. L. R., Huk, A. C., & Pillow, J. W. (2015). Single-trial spike trains in parietal cortex reveal discrete steps during decision-making. *Science, 349*, 184–187.

Lau, H. (2019). Consciousness, metacognition, and perceptual reality monitoring. Retrieved from https://psyarxiv.com/ckbyf/.

Lee, S. H., Blake, R., & Heeger, D. J. (2005). Traveling waves of activity in primary visual cortex during binocular rivalry. *Nature Neuroscience, 8*, 22–23.

Lehky, S. R., & Sejnowski, T. J. (1999). Seeing white: Qualia in the context of decoding population codes. *Neural Computation, 11*, 1261–1280.

Ma, W. J. (2012). Organizing probabilistic models of perception. *Trends in Cognitive Sciences, 16*, 511–518.

Ma, W. J., Beck, J. M., Latham, P. E., & Pouget, A. (2006). Bayesian inference with probabilistic population codes. *Nature Neuroscience, 9*, 1432–1438.

Ma, W. J., & Jazayeri, M. (2014). Neural coding of uncertainty and probability. *Annual Review of Neuroscience, 37*, 205–220.

Maloney, L. T., & Knoblauch, K. (2020). Measuring and modeling visual appearance. *Annual Review of Vision Science, 6*, 13.1–13.19.

Mamassian, P. (2016). Visual Confidence. *Annual Review of Vision Science, 2*, 459–481.

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. San Francisco, CA: W. H. Freeman.

Mazor, M., Friston, K. J., & Fleming, S. M. (2020). Distinct neural contributions to metacognition for detecting, but not discriminating visual stimuli. *eLife, 9*, e53900.

Meuwese, J. D. I., van Loon, A. M., Lamme, V. A. F., & Fahrenfort, J. J. (2014). The subjective experience of object recognition: Comparing metacognition for object detection and object categorization. *Attention, Perception, and Psychophysics, 76*, 1057–1068.

Meyniel, F., Sigman, M., & Mainen, Z. F. (2015). Confidence as Bayesian probability: From neural origins to behavior. *Neuron, 88*, 78–92.

Moreno-Bote, R., Knill, D. C., & Pouget, A. (2011). Bayesian sampling in visual perception. *Proceedings of the National Academy of Sciences of the United States of America, 108*, 12491–12496.

Morrison, J. (2016). Perceptual confidence. *Analytic Philosophy, 57*, 15–48.

Morrison, J. (2017). Perceptual confidence and categorization. *Analytic Philosophy, 58*, 71–85.

Moutard, C., Dehaene, S., & Malach, R. (2015). Spontaneous fluctuations and nonlinear ignitions: Two dynamic faces of cortical recurrent loops. *Neuron, 88*, 194–206.

Munton, J. (2016). Visual confidences and direct perceptual justification. *Philosophical Topics, 44*, 301–326.

Nanay, B. (2020). Perceiving indeterminately. *Thought: A Journal of Philosophy, 9*(3), 160–166.

Noy, N., Bickel, S., Zion-Golumbic, E., Harel, M., Golan, T., Davidesco, I., . . . Malach, R. (2015). Ignition's glow: Ultra-fast spread of global cortical activity accompanying local "ignitions" in visual cortex during conscious visual perception. *Consciousness Cognition, 35*, 206–224.

O'Connell, R. G., Dockree, P. M., & Kelly, S. P. (2012). A supramodal accumulation-to-bound signal that determines perceptual decisions in humans. *Nature Neuroscience, 15*, 1729–1735.

Olshausen, B. A., & Field, D. J. (1996). Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature, 381*, 607–609.

Orbán, G., Berkes, P., Fiser, J., & Lengyel, M. (2016). Neural variability and sampling-based probabilistic representations in the visual cortex. *Neuron, 92*, 530–543.

Ott, T., Masset, P., & Kepecs, A. (2019). The neurobiology of confidence: From beliefs to neurons. *Cold Spring Harbor Symposia on Quantitative Biology, 83*, 038794.

Overgaard, M., Fehl, K., Mouridsen, K., Bergholt, B., & Cleeremans, A. (2008). Seeing without seeing? Degraded conscious vision in a blindsight patient. *PLoS One, 3*, e3028.

Pereira, M., Mégevand, P., Tan, M. X., Chang, W., Wang, S., Rezai, A., . . . Faivre, N. (2020). Evidence accumulation determines conscious access. *bioRxiv*. Advance online publication. doi:10.1101/2020.07.10.196659.

Persaud, N., McLeod, P., & Cowey, A. (2007). Post-decision wagering objectively measures awareness. *Nature Neuroscience, 10*, 257–261.

Petrovici, M. A., Bill, J., Bytschok, I., Schemmel, J., & Meier, K. (2016). Stochastic inference with spiking neurons in the high-conductance state. *Physical Review E, 94*, 042312.

Polonsky, A., Blake, R., Braun, J., & Heeger, D. J. (2000). Neuronal activity in human primary visual cortex correlates with perception during binocular rivalry. *Nature Neuroscience, 3*, 1153–1159.

Pouget, A., Drugowitsch, J., & Kepecs, A. (2016). Confidence and certainty: Distinct probabilistic quantities for different goals. *Nature Neuroscience, 19*, 366–374.

Rahnev, D. (2017). The case against full probability distributions in perceptual decision making. *bioRxiv*. Advance online publication. doi:10.1101/108944.

Rahnev, D., & Denison, R. N. (2018). Suboptimality in perceptual decision making. *Behavioral and Brain Sciences, 41*, 1–107.

Ratcliff, R., Smith, P. L., Brown, S. D., & McKoon, G. (2016). Diffusion decision model: Current issues and history. *Cognitive Science, 20*, 260–281.

Ress, D., & Heeger, D. J. (2003). Neuronal correlates of perception in early visual cortex. *Nature Neuroscience, 6*, 414–420.

Sanborn, A. N., & Chater, N. (2016). Bayesian brains without probabilities. *Trends in Cognitive Science, 20*, 883–893.

Sanger, T. D. (1996). Probability density estimation for the interpretation of neural population codes. *Journal of Neurophysiology, 76*, 2790–2793.

Savin, C., & Denève, S. (2014). Spatio-temporal representations of uncertainty in spiking neural networks. *Advances in Neural Information Processing Systems, 27*, 2024–2032.

Sergent, C., & Dehaene, S. (2004). Is consciousness a gradual phenomenon? Evidence for an all-or-none bifurcation during the attentional blink. *Psychological Science, 15*, 720–728.

Shadlen, M. N., & Kiani, R. (2011). Consciousness as a decision to engage. In S. Dehaene & Y. Christen (Eds.), *Characterizing consciousness: From cognition to the clinic?* (Vol. 31, pp. 27–46). Berlin: Springer.

Shivkumar, S., Lange, R. D., Chattoraj, A., & Haefner, R. M. (2018). A probabilistic population code based on neural samples. Retrieved from https://arxiv.org/abs/1811.09739.

Shoemaker, S. (2007). *Physical realization*. Oxford: Oxford University Press.

Sundareswara, R., & Schrater, P. R. (2008). Perceptual multistability predicted by search model for Bayesian decisions. *Journal of Vision, 8*, 12.11–19.

Tagliabue, C. F., Veniero, D., Benwell, C. S. Y., Cecere, R., Savazzi, S., & Thut, G. (2019). The EEG signature of sensory evidence accumulation during decision formation closely tracks subjective perceptual experience. *Scientific Reports, 9*, 4949.

Trommershäuser, J., Kording, K., & Landy, M. S., (Eds.). (2011). *Sensory cue integration*. Oxford: Oxford University Press.

Twomey, D. M., Murphy, P. R., Kelly, S. P., & O'Connell, R. G. (2015). The classic P300 encodes a build-to-threshold decision variable. *European Journal of Neuroscience, 42*, 1636–1643.

van Bergen, R. S., Ma, W. J., Pratte, M. S., & Jehee, J. F. M. (2015). Sensory uncertainty decoded from visual cortex predicts behavior. *Nature Neuroscience, 18*, 1728–1730.

van Vugt, B., Dagnino, B., Vartak, D., Safaai, H., Panzeri, S., Dehaene, S., & Roelfsema, P. R. (2018). The threshold for conscious report: Signal loss and response bias in visual and frontal cortex. *Science, 23*, eaar7186.

Vul, E., Hanus, D., & Kanwisher, N. (2009). Attention as inference: Selection is probabilistic; responses are all-or-none samples. *Journal of Experimental Psychology: General, 138*, 546–560.

Walker, E. Y., Cotton, R. J., Ma, W. J., & Tolias, A. S. (2020). A neural basis of probabilistic computation in visual cortex. *Nature Neuroscience, 23*, 122–129.

Weiskrantz, L. (1996). Blindsight revisited. *Current Opinion Neurobiology, 6*, 215–220.

Wixted, J. T. (2019). The forgotten history of signal detection theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 46*, 201–233.

Wunderlich, K., Schneider, K. A., & Kastner, S. (2005). Neural correlates of binocular rivalry in the human lateral geniculate nucleus. *Nature Neuroscience, 8*, 1595–1602.

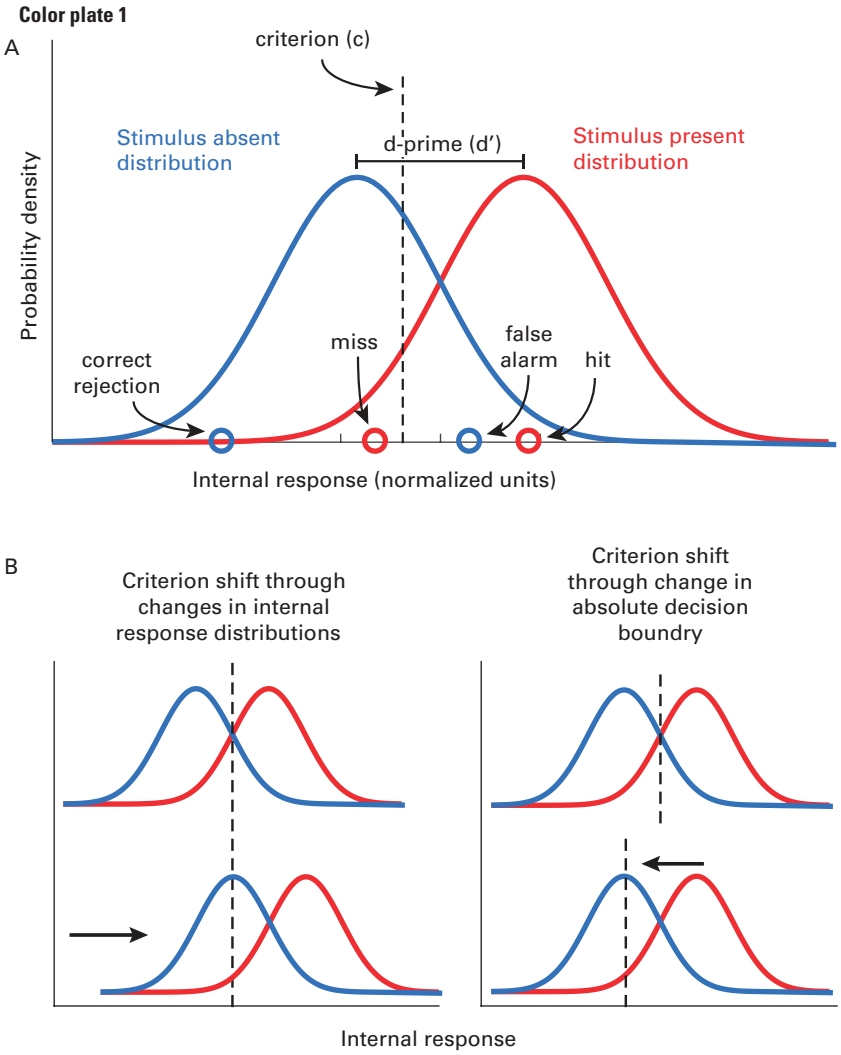
Yang, T., & Shadlen, M. N. (2007). Probabilistic reasoning by neurons. *Nature, 447*, 1075–1080.

Yeon, J., & Rahnev, D. (2020). The suboptimality of perceptual decision making with multiple alternatives. *Nature Communications,* 11, 3857.

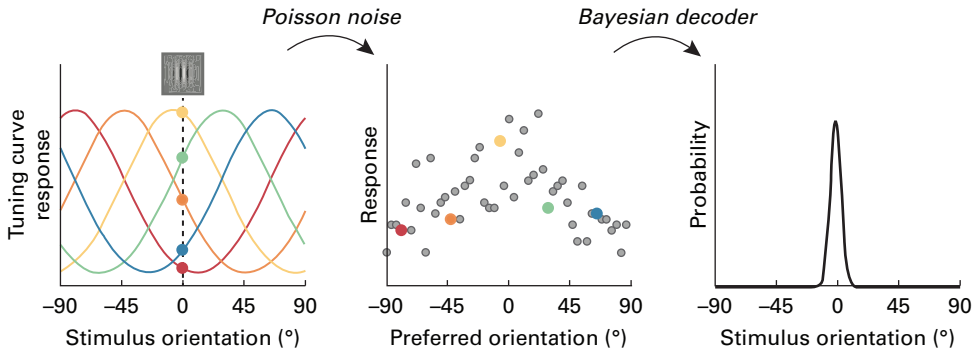
Zemel, R. S., Dayan, P., & Pouget, A. (1998). Probabilistic interpretation of population codes. *Neural Computation, 10*, 403–430.

Zhou, Y., & Freedman, D. J. (2019). Posterior parietal cortex plays a causal role in perceptual and categorical decisions. *Science, 365*, 180–185.

Color plate 1

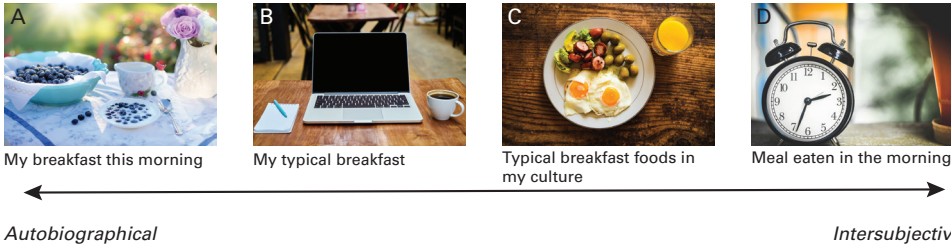


Color plate 1 (previous page)
Signal detection theory (SDT) illustrated for a detection task. (A) The sensory stage of SDT is a single internal response, generated on every trial (open circles on the x-axis represent four example trials). When no stimulus is presented, the internal response is drawn from a Gaussian probability distribution (stimulus-absent distribution). When the stimulus is presented, the internal response is drawn from another Gaussian distribution with a higher mean (stimulus-present distribution). When the internal responses are normalized by the standard deviation of the Gaussian, the discriminability of the two distributions (d') is the difference in distribution means. The criterion represents the decision stage and is independent of d'. It defines the magnitude of internal response needed to report that the stimulus was present. When the stimulus is absent, an internal response below the criterion (report absent) gives a correct rejection, and an internal response above the criterion (report present) gives a false alarm (open circles). When the stimulus is present, an internal response below the criterion gives a miss, and an internal response above gives a hit (open circles). (B) The SDT criterion computed from behavior is a relative measure: it indicates how much evidence is needed to make a decision relative to the intersection point of the two distributions. However, when relating behavior to neural activity, we may be interested in the absolute decision boundary—for example, what neural response magnitude is required to report that the stimulus was present. This example illustrates a potential difficulty in inferring an absolute decision boundary from the SDT criterion in a detection task. The criterion estimated from behavior could change due to either a shift in the internal response distributions with no change in absolute decision boundary (left) or a shift in the absolute decision boundary with no change in internal response distributions (right).



Color plate 2

Probabilistic population codes. Left: A population of orientation-tuned neurons has tuning curves that tile orientation. Example tuning curves from the population are shown (curves), along with their mean responses to a stimulus with 0° (i.e., vertical) orientation (points). Middle: On a single trial, the response of each neuron (gray dots) is determined by its tuning curve and Poisson noise; the preferred orientation of each neuron is the orientation at which its tuning curve peaks. Right: A probability distribution can be read out from the single-trial population response using a Bayesian decoder, which combines the information from all neurons to calculate the likelihood of the population response given a stimulus of each orientation.



Color plate 3

Hypothesized placement of examples of different kinds of memories in terms of their level of autobiographicality versus intersubjectivity/sharedness. (Images are open-source with free license.)

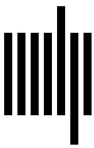
Neuroscience and Philosophy

Edited by: Felipe De Brigard, Walter Sinnott-Armstrong

Citation:

Neuroscience and Philosophy
Edited by: Felipe De Brigard, Walter Sinnott-Armstrong
DOI: 10.7551/mitpress/12611.001.0001
ISBN (electronic): 9780262367332
Publisher: The MIT Press
Published: 2022

The open access edition of this book was made possible by generous funding and support from MIT Press Direct to Open



The MIT Press

The MIT Press would like to thank the anonymous peer reviewers who provided comments on drafts of this book. The generous work of academic experts is essential for establishing the authority and quality of our publications. We acknowledge with gratitude the contributions of these otherwise uncredited readers.

This book was set in Stone Serif by Westchester Publishing Services. .

Library of Congress Cataloging-in-Publication Data

Names: Brigard, Felipe de, editor. | Sinnott-Armstrong, Walter, 1955– editor.
Title: Neuroscience and philosophy / edited by Felipe De Brigard and Walter Sinnott-Armstrong.
Description: Cambridge, Massachusetts : The MIT Press, [2022] | Includes bibliographical references and index.
Identifiers: LCCN 2021000758 | ISBN 9780262045438 (paperback)
Subjects: LCSH: Cognitive neuroscience—Philosophy.
Classification: LCC QP360.5 .N4973 2022 | DDC 612.8/233—dc23
LC record available at https://lccn.loc.gov/2021000758