# Multistability and Perceptual Inference

**Samuel J. Gershman**
*sjgershm@princeton.edu*
*Department of Psychology and Neuroscience Institute, Princeton University,*
*Princeton, NJ, U.S.A.*

**Edward Vul**
*evul@ucsd.edu*
*Department of Psychology, University of California, San Diego, CA 92093, U.S.A.*

**Joshua B. Tenenbaum**
*jbt@mit.edu*
*Department of Brain and Cognitive Sciences, MIT, Cambridge, MA 02139, U.S.A.*

**Ambiguous images present a challenge to the visual system: How can uncertainty about the causes of visual inputs be represented when there are multiple equally plausible causes? A Bayesian ideal observer should represent uncertainty in the form of a posterior probability distribution over causes. However, in many real-world situations, computing this distribution is intractable and requires some form of approximation. We argue that the visual system approximates the posterior over underlying causes with a set of samples and that this approximation strategy produces perceptual multistability—stochastic alternation between percepts in consciousness. Under our analysis, multistability arises from a dynamic sample-generating process that explores the posterior through stochastic diffusion, implementing a rational form of approximate Bayesian inference known as Markov chain Monte Carlo (MCMC). We examine in detail the most extensively studied form of multistability, binocular rivalry, showing how a variety of experimental phenomena—gamma-like stochastic switching, patchy percepts, fusion, and traveling waves—can be understood in terms of MCMC sampling over simple graphical models of the underlying perceptual tasks. We conjecture that the stochastic nature of spiking neurons may lend itself to implementing sample-based posterior approximations in the brain.**

## 1 Introduction

We perceive the world through our senses, but any stream of sense data is a fundamentally impoverished source of information about the external world. To take the most familiar example, retinal images are

two-dimensional, while the external world is three-dimensional. Our brains must go beyond what is directly available in the sensory data through a process of interpretation to achieve the rich percepts of conscious awareness. Bayesian statistics has provided a powerful framework for understanding this inferential process (Knill & Richards, 1996); its central postulate is that percepts represent hypotheses about the external world whose degree of belief (the posterior probability) is determined by a combination of sensory evidence (the likelihood) and background assumptions about a priori plausible world structures (the prior). Numerous experiments have confirmed the explanatory reach of Bayesian models in perception as well as cognition, motor control, and other areas of brain function (Brainard & Freeman, 1997; Weiss, Simoncelli, & Adelson, 2002; Kording & Wolpert, 2004; Griffiths & Tenenbaum, 2006). Computer vision and artificial intelligence researchers have shown how Bayesian inference operating over networks of random variables known as probabilistic graphical models can solve many real-world perceptual inference and higher-level reasoning tasks (Geman & Geman, 1984; Koller & Friedman, 2009). Computational neuroscientists have formulated hypotheses about how populations of neurons can encode, transform, and decode posterior distributions (Knill & Pouget, 2004).

Despite these successes arguing for a unifying Bayesian view of the brain, a major computational obstacle remains: Bayesian inference in large-scale graphical models needed to capture real-world perceptual tasks is computationally intractable. Posterior degrees of belief in these models cannot, in general, be calculated exactly by any resource-limited computational device. Bayesian statisticians and researchers in computer vision and artificial intelligence (AI) routinely use a range of approaches and algorithms for approximating ideal Bayesian inference in graphical models (Koller & Friedman, 2009). Thus, Bayesian theories of perceptual inference remain incomplete until they can explain how the brain approximates posterior inference effectively and quickly.

Early Bayesian models of human perception proposed that instead of calculating the full posterior distribution, the brain finds only the point estimate with highest posterior probability (the maximum a posteriori, or MAP, estimate). Although MAP estimation provides a remarkably good account of many experimental phenomena (Schrater & Kersten, 2000; Weiss et al., 2002), it is not sufficient to account for human behavior. Not only is MAP estimation suboptimal because it fails to represent uncertainty about perceptual interpretations, but more important, it can not account for recent evidence that humans and other animals represent their uncertainty in perceptual tasks (Grinband, Hirsch, & Ferrera, 2006; Kiani & Shadlen, 2009). Thus, we must consider alternative approximate inference strategies the brain might adopt.

Here we argue that perceptual multistability—alternation between conscious percepts in response to ambiguous sensory inputs (Blake &

Logothetis, 2002)—provides both qualitative and quantitative evidence for a different and more powerful approach to approximate Bayesian inference in the brain, which can be formalized in terms of algorithms from Bayesian statistics, computer vision, and AI. Specifically, we show how perceptual multistability can be naturally explained as a consequence of Monte Carlo inference, approximating Bayesian posterior distributions with a set of samples (Robert & Casella, 2004). An emerging literature suggests that Monte Carlo approximations might be employed in human (Sanborn, Griffiths, & Navarro, 2006; Levy, Reali, & Griffiths, 2009; Vul, Goodman, Griffiths, & Tenenbaum, 2009) and animal (Daw & Courville, 2007) cognition, and several authors have proposed that the spiking activity of neurons might be interpreted as a set of samples (Hoyer & Hyvärinen, 2003; Lee & Mumford, 2003; Shi & Griffiths, 2009; Fiser, Berkes, Orbán, & Lengyel, 2010). We suggest that Markov chain Monte Carlo (MCMC) methods are particularly well suited to modeling perceptual multistability. These methods provide state-of-the-art engineering performance on the large-scale graphical models needed in computer vision. Moreover, when MCMC algorithms are applied to graphical models, they produce spatiotemporal dynamics in inference that correspond to experimentally observed phenomena of multistability, as we explore below. Prior work has argued that stochastic sampling produces perceptual multistability (Schrater & Sundareswara, 2007; Sundareswara & Schrater, 2008) in the case of the Necker cube. We go beyond this work by showing that a rational MCMC algorithm for approximate Bayesian inference in graphical models produces a broad range of perceptual multistability phenomena while approximating a rational solution to a real vision problem: inferring the correct image interpretation from noisy binocular inputs.

We focus on the most extensively studied form of perceptual multistability, binocular rivalry, in which incompatible images (e.g., orthogonal gratings) are presented dichoptically to the eyes. Under most (but not all) conditions, this results in a percept that switches stochastically between the two images (see Blake, 2001, for a review). We first formulate a simple graphical model for retinal images, such that inference in this model using Bayes's rule recovers the latent scene underlying sensory inputs. We then show how approximate Bayesian inference in this model using MCMC can reproduce a variety of psychophysical phenomena. Finally, we discuss the connections between MCMC computations and possible neural implementations.

Our model is at present a caricature of the visual processing pathway that gives rise to multistable perception, and binocular rivalry in particular. We have endeavored to simulate only a handful of the numerous phenomena for which quite sophisticated and neurally realistic models have been developed (Laing & Chow, 2002; Freeman, 2005; Moreno-Bote, Rinzel, & Rubin, 2007; Noest, van Ee, Nijs, & Van Wezel, 2007; Wilson, 2007). Our goal in this article is not to supplant the existing theoretical work but rather
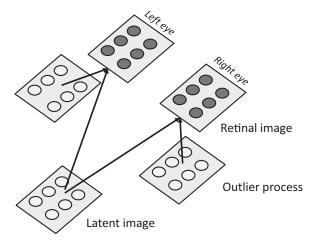
Figure 1: Generative model. Schematic illustrating the probabilistic process by which retinal images are generated. Shaded nodes denote observed variables; unshaded nodes denote unobserved (latent) variables. Arrows denote causal dependencies.

to supplement it by providing a normative statistical motivation for the dynamical processes posited by these models. We believe these ideas are best illustrated in a simplified setting, in which their mechanistic principles are relatively transparent. The idea of sampling over a graphical model structure is very general and can be applied to more sophisticated versions of the model we present.

## 2 A Probabilistic Model of Binocular Rivalry

Our starting point is a probabilistic generative model of retinal images that represents the brain's assumptions about how its sensory inputs were generated (see Figure 1). Formally, this model is a variant of a Markov Random Field (MRF), one of the most standard forms of probabilistic graphical model used in computer vision (Geman & Geman, 1984). In the appendix, we describe this model mathematically, but here we focus on the basic intuitions behind this model. The model posits two sets of latent variables: a latent scene and an eye-specific outlier process that governs whether a given patch of retina observes the latent scene. The goal of the visual system is to infer the underlying image given the observed retinal input.

We represent the latent scene and the two retinal images as arrays of luminance values (e.g., a $32 \times 32$ grayscale pixel image), while the outlier processes are represented as binary arrays of the same size. The luminance values for the pixels in each eye are the noise-perturbed luminance of the

corresponding pixels of the scene. However, if a pixel from one eye is deemed to be an outlier, then the luminance at that pixel is meaningless and contains no information with respect to the scene. As we discuss below, this scene representation is admittedly highly simplified, but it is adequate for our goal of studying the dynamics of inference in multistable perception rather than scene perception per se.

This outlier process allows the possibility that sometimes the retinal signal does not correspond to the scene at all—for instance, if an occluder blocks a portion of one eye's field of view (Shimojo & Nakayama, 1990) or if a portion of the cornea or retina of one eye is damaged. Thus, the outlier process captures the notion that if, for one of these reasons, a portion of the retinal image does not correspond to the scene, it should simply be ignored (suppressed). This aspect of our model allows the two retinal images to compete in rivalry for representation in the inferred scene: if the two eyes see conflicting images, at least one of them must not represent the actual scene present in the world.

The psychological and neural plausibility of the outlier process is motivated by evidence that suppression during binocular rivalry is accompanied by a general loss in visual sensitivity (Blake, 2001). For example, target detection performance is lower (Wales & Fox, 1970; Fukuda, 1981) and reaction times are longer (Fox & Check, 1968; O'Shea, 1987) for probe stimuli presented to the suppressed eye compared to probe stimuli presented to the dominant eye. Moreover, single-unit recordings (Leopold & Logothetis, 1996) and functional magnetic resonance imaging (fMRI) studies (Polonsky, Blake, Braun, & Heeger, 2000) have demonstrated inhibition of responses in primary visual cortex during suppression periods. These findings indicate that the visual system is engaged not only in inferring the latent causes of its sensory inputs, but also inferring the reliability of its measurement devices.

Crucially, this outlier process is spatially smooth, so that two adjacent pixels are more likely to have the same outlier status. This corresponds to the idea that if a given portion of retina is not receiving a signal corresponding to the scene, its neighbors are also not likely to represent the scene, since occlusion, macular degeneration, retinitis pigmentosa, cataracts, or something else disrupts the field of view in a spatially smooth manner.

Thus far, we have described a probabilistic graphical model designed to infer the luminance value of each pixel of an underlying two-dimensional latent scene independently. However, the visual system has a much more ambitious goal: inferring a complete, semantically coherent, three-dimensional latent scene. A full generative model of the scenes that the visual system considers would involve many complex constraints and incompletely understood assumptions; developing such a model would be hard to do without a complete account of human vision and visual cognition. Instead, as a proxy for such a rich model, we assume an image-like scene representation with luminance values that correspond to some combination of the luminance values of the real images that comprise the stimulus and that the

degree to which pixels correspond to one stimulus or another is spatially smooth. This smooth "latent image" prior is a simple way to encode the idea of spatially organized semantic coherence in the inferred latent scene.

Specifically, we approximate a rich hypothesis space of latent images without having to specify a complete distribution over natural images, by assuming that each latent scene pixel $s_n$ is a linear combination of the corresponding real images pixels ($x_n^L$ and $x_n^R$) in the stimulus: $s_n = w_n x_n^L + (1 - w_n)x_n^R$, where $w_n$ can theoretically range over the entire real line and is assumed to be spatially smooth. This can be understood as an "empirical" approximation to the distribution over natural images, using only linear combinations of the observed images. In practice, this heuristic will work nearly as well, since the posterior distribution will be strongly concentrated around the two retinal images.

## 3 Approximate Inference by Sampling

Given an observed pair of retinal images, Bayes's rule can be used to invert the generative model and recover the latent scene:

$$P(\mathbf{s}|\mathbf{x}) = \frac{1}{Z}P(\mathbf{s}) \sum_{\pi} P(\mathbf{x}|\mathbf{s}, \pi)P(\pi), \tag{3.1}$$

where $Z$ is a normalizing constant and $\pi$ denotes the outlier process. Because it is generally intractable to calculate the normalizing constant, approximations must be considered. Monte Carlo methods (Robert & Casella, 2004) approximate the posterior with a set of $M$ samples:

$$P(\mathbf{s}|\mathbf{x}) \approx \frac{1}{M} \sum_{m=1}^{M} \delta\big[\mathbf{s}^{(m)}, \mathbf{s}\big], \tag{3.2}$$

where $\delta[\cdot, \cdot]$ is the Kronecker delta function whose value is 1 when its arguments are equal, and 0 otherwise. The challenge is to draw samples from the posterior; in most cases, the posterior cannot be sampled from directly. The key idea behind MCMC methods is that a sequence of distributions (comprising a Markov chain) can be constructed that eventually converges to the posterior. This means that after a burn-in period, samples from this Markov chain will be distributed according to the posterior. Gibbs sampling (Geman & Geman, 1984), perhaps the best-known MCMC method, was originally developed for MRFs similar to our model. It proceeds by performing sweeps over the latent variables (scene and outlier nodes), sampling each variable from its conditional distribution while holding all the other variables fixed (see the appendix for details). In our implementation, the order in which nodes are updated in each sweep is randomized. Due to the topology of the MRF, each conditional distribution depends on only a

subset of the latent variables (see the appendix), and hence Gibbs sampling can be done using only information that is local to each latent variable.

## 4 Results

We now illustrate the behavior of our sampling model using several key paradigms from the binocular rivalry literature. Our goal is to highlight the conceptual principle underlying our explanation of multistable perception: stochastic perceptual dynamics arise from approximate inference in a graphical model of visual inputs. Details about parameter selection and robustness can be found in the appendix.

**4.1 Dynamics of Perceptual Switches.** In a typical binocular rivalry experiment, subjects are asked to say which of two images corresponds to their global percept. To make the same query of the current state of our simulated Markov chain, we defined a perceptual switch to occur when at least 9/10 of the scene nodes take on the same value. Figure 2A shows a sample time course. It may seem surprising that the model spends relatively little time near the extremes and that switches are fairly gradual. This phenomenonology is consistent with several experiments showing that substantial time is spent in transition periods (Mueller & Blake, 1989; Brascamp, van Ee, Noest, Jacobs, & van den Berg, 2006; Kang, 2009).

One of the most robust findings in the literature on binocular rivalry is that switching times between different stable percepts tend to follow a gamma-like distribution (Blake, 2001). In other words, the dominance durations of stability in one mode tend to be neither overwhelmingly short nor long. Figure 2B shows the histogram of dominance durations and the maximum-likelihood fit of a gamma distribution to our simulations and real data, demonstrating that the durations produced by MCMC are well described by a gamma distribution.

To account for this gamma-like switching behavior, many papers have described neural circuits that could produce switching oscillations with the right stochastic dynamics by suggesting a competition between mutual inhibition and adaptation (Moreno-Bote et al., 2007; Wilson, 2007). Similarly, existing rational process models of multistability (Dayan, 1998; Schrater & Sundareswara, 2007; Sundareswara & Schrater, 2008) add specific adaptation- or memory-like constraints to produce this effect. In our model, these additional memory constraints are unnecessary because the spatial coupling of adjacent nodes in the latent scene creates a form of memory in the inference.

The gamma distribution arises in MCMC on an MRF because each hidden node takes an approximately exponentially distributed amount of time to switch, but these switches must co-occur in sequence. So the total amount of time until enough nodes switch to one mode will be the sum of exponential random variables, or a gamma distribution. Thus, gamma-distributed
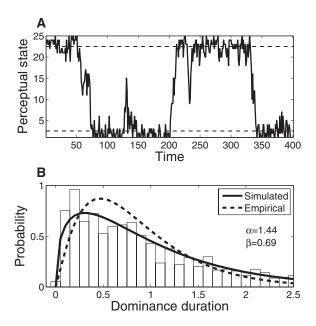
Figure 2: Perceptual switching dynamics. (A) Simulated time course of bistability. Plotted on the *y*-axis is the number of nodes with value greater than 0.5. The horizontal lines show the thresholds for a perceptual switch. (B) Distribution of simulated dominance durations (mean-normalized) for MRF with lattice topology. Curves show gamma distributions fitted to simulated (with parameter values shown on the right) and empirical data, replotted from Mamassian and Goutcher (2005).

dominance durations fall naturally out of MCMC operating on a spatially smooth MRF.

**4.2  Piecemeal Rivalry and Traveling Waves.** Another empirical observation about spatiotemporal dynamics in rivalry is that stability is often incomplete across the visual field, producing piecemeal rivalry, in which one portion of the visual field looks like the image in one eye, while another portion looks like the image in the other eye (Mueller & Blake, 1989). One intriguing feature of these piecemeal percepts is the phenomenon known as traveling waves: subjects tend to perceive a perceptual switch as a wave propagating over the visual field (Wilson, Blake, & Lee, 2001; Lee, Blake, & Heeger, 2005): the suppressed stimulus becomes dominant in an isolated location of the visual field and then gradually spreads. These traveling waves reveal an interesting local dynamics during an individual switch itself.

Demonstrating the dynamics of traveling waves within patches of the percept requires a different method of probing perception instead of asking
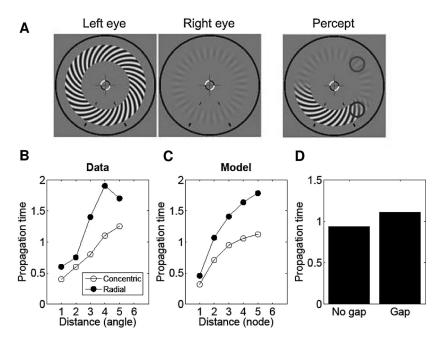
Figure 3: Traveling waves. (A) Annular stimuli used by Lee et al. (2005) (left and center panels) and the subject percept reported by observers (right panel), in which the low-contrast stimulus was seen to spread around the annulus, starting at the top. Figure reprinted with permission from Lee et al. (2005). (B) Propagation time as a function of distance around the annulus, replotted from Wilson et al. (2001). Filled circles represent radial gratings, and open circles represent concentric gratings. A transient increase in contrast of the suppressed stimulus induces a perceptual switch at the location of contrast change. The propagation time for a switch at a probe location increases with distance (around the annulus) from the switch origin. (C) Simulated propagation time (measured by the time to switch percept following a switch at varying distances around the annulus). (D) Average simulated propagation time between nodes separated by a gap compared to nodes without a gap.

subjects to evaluate the "global percept." Wilson et al. (2001) used annular stimuli (see Figure 3A) and probed a particular patch along the annulus. They showed that the time at which the suppressed stimulus in the test patch becomes dominant is a function of the distance (around the circumference of the annulus) between the test patch and the patch where a dominance switch was induced by transiently increasing the contrast of the suppressed stimulus (see Figure 3B). This dependence of switch time on distance suggested to Wilson et al. that stimulus dominance was propagating around the annulus. Wilson et al. made two additional observations about the nature
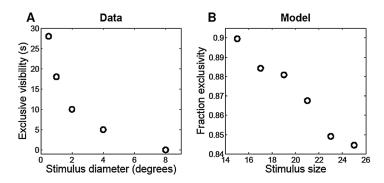
Figure 4: Piecemeal rivalry. (A) Exclusive visibility as a function of stimulus size for a single subject. Replotted from O'Shea et al. (1997). (B) Simulated fraction exclusivility as a function of stimulus size.

of traveling waves: (1) propagation is faster along a collinear (concentric) contour compared to an orthogonal (radial) contour, and (2) introducing a gap in the annulus substantially slows or halts propagation. Subsequently, using fMRI, Lee et al. (2005) showed that the propagation of this traveling wave can be observed in primary visual cortex.

To simulate such traveling waves within the percept of a stimulus, we constructed a $7 \times 7$ MRF with annular topology and measured the propagation time at different nodes along the annulus. We computed pairwise correlation between nodes and converted this to propagation time by exponentially transforming the negative correlation and then rescaling it to match the empirical propagation times. Figure 3C shows the simulated propagation time between image nodes increasing as a function of distance around the annulus. Since the spatial coupling induces local dependencies in the MRF, nodes will be more likely to switch once their neighbors have switched, thus producing a domino effect around the ring. We can account for faster propagation of traveling waves in the concentric rings condition by increasing the strength of spatial coupling for such collinear stimuli (following the observed greater coupling of dominance in collinear patches; see Alais & Blake, 1999). Moreover, since a missing node disrupts the domino effect of spatial coupling, we can reproduce the disruptive effect of a gap on traveling wave propagation (Wilson et al., 2001) by removing one of the nodes in the annulus and comparing propagation with and without the gap (see Figure 3D).

Another factor influencing piecemeal rivalry is the size of the rivaling stimuli. Several experiments have reported that exclusivity (the amount of time spent in a stable percept) decreases as a function of stimulus size (see Figure 4A; Blake, O'Shea, & Mueller, 1992; O'Shea, Sims, & Govan, 1997). We simulated this finding by manipulating the proportion of the retinal images

occupied by the stimuli. Figure 4B shows the results of this simulation, consistent with the experimental findings. The intuitive explanation for this phenomenon is that larger stimuli require more individual node switches to achieve stable percepts; thus, there will be longer periods in which the scene configuration occupies local minima of the energy landscape corresponding to piecemeal percepts.

**4.3  Binocular Fusion.**  In addition to the mode-hopping behavior that characterizes binocular rivalry, bistable percepts often produce other states. In some conditions, the two percepts are known to fuse rather than rival: the percept then becomes a composite or superposition of the two stimuli (Brascamp et al., 2006). This fused perceptual state can be induced most reliably by decreasing the distance in feature space between the two stimuli (Knapen, Kanai, Brascamp, van Boxtel, & van Ee, 2007) or decreasing the contrast of both stimuli (Liu, Tyler, & Schor, 1992; Burke, Alais, & Wenderoth, 1999).

Fusion is documented in experiments where subjects are given three options to report their percept: one of two global percepts or something in between. We define such a fused percept as a perceptual state lying between the two bistable modes—that is, an interpretation between the two rivalrous, high-probability interpretations. On our account, manipulation of distance in feature space amounts to varying the distance between the two modes, and reduction of contrast corresponds to an increase in the variance around the modes. By making the modes closer together or increasing the variance parameter ($\sigma_i^2$) in the model, greater probability mass is assigned to the intermediate interpretation—a fused percept. Thus, these manipulations shift the energy landscape so as to systematically increase the odds of fused percepts, matching the phenomenology of these stimuli (see Figure 5).

**4.4  Contrast Dependence of Dominance Durations.**  A classic observation in the psychophysics of binocular rivalry is that when the contrast in one eye (the variable-contrast eye) is increased, the dominance durations in the other (fixed-contrast) eye decrease while the dominance durations in the variable-contrast eye stay relatively unchanged. This phenomenon is often referred to as Levelt's second proposition (Levelt, 1965) and has been a key target for many models of binocular rivalry (Dayan, 1998; Lankheet, 2006; Moreno-Bote et al., 2007; Wilson, 2007). A standard explanation of this phenomenon appeals to the interaction between two dynamic processes: short-timescale changes in mutual inhibition between monocular neural populations and long-timescale changes in adaptation. The neurons representing the currently dominant stimulus inhibit the neurons representing the currently suppressed stimulus, but adaptation of the inhibiting neurons causes this suppression to wane over time, allowing the previously
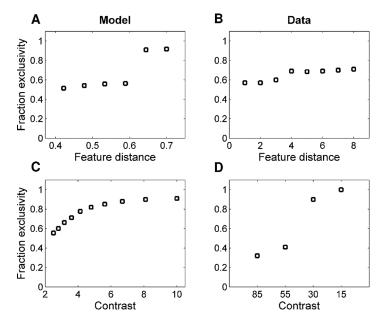
Figure 5: Fusion of binocular images. Simulated (A) and empirical (B) fraction of exclusive percepts as a function of feature space distance between the binocular stimuli. Data replotted from Knapen et al. (2007). Simulated (C) and empirical (D) fraction of exclusive percepts as a function of stimulus contrast. Model contrast is measured by the inverse of the variance parameter, $\sigma_i^2$. Data replotted from Burke et al. (1999).

inhibited neurons to accumulate enough activation to cause a perceptual switch.

Brascamp et al. (2006) have observed that Levelt's second proposition is valid only when the stimulus presented to the fixed-contrast eye is relatively high contrast. Under low-contrast conditions, the proposition actually reverses (see Figure 6). Brascamp et al. (2006) suggested amending Levelt's second proposition to: "Changes to one of the two contrasts mainly affect dominance durations in the higher contrast eye."

Our model captures the contrast dependence of dominance durations under both high and low fixed-contrast conditions (see Figure 6). On our account, this effect arises from transitions between several inference regimes:

1. When contrast to both eyes is (equally) low, then the data in both eyes are uncertain and easy to discount; therefore, the modes corresponding to one or another eye being dominant are easy to escape, fused percepts are common (see section 4.3), and dominance durations in both eyes will be brief.
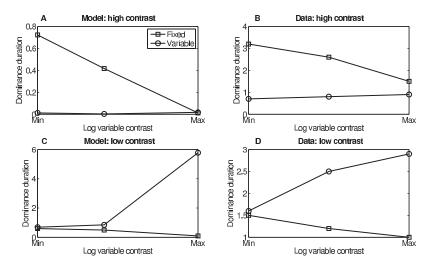
Figure 6: Effect of stimulus contrast on dominance durations. Simulated (A) and experimental (B) dominance durations as a function of stimulus contrast in the variable contrast eye (*x*-axis), when the stimulus presented to the fixed-contrast eye is high contrast (the maximum of the contrast range presented to the variable contrast eye). Simulated (C) and experimental (D) dominance durations as a function of stimulus contrast in the variable-contrast eye when the stimulus presented to the fixed-contrast eye is low contrast (the minimum of the contrast range presented to the variable contrast eye). Data replotted from Brascamp et al. (2006).

2. When contrast is low in one eye but high in the other, the low-contrast mode is easy to escape; moreover, the high-contrast eye is hard to discount as an outlier in favor of the low-contrast eye. Altogether, the higher-contrast eye enjoys long dominance durations.
3. When the contrast in both eyes is high, fused or mixed percepts are unlikely and do not last long (Hollins, 1980). Moreover, neither eye is easier to discount: therefore, the advantage that the relatively high-contrast eye had in regime 2 no longer exists, and dominance durations in both eyes will again be brief.

Thus, when the fixed-contrast eye has high contrast and the contrast of the variable-contrast eye is increased to match it, the system transitions from regime 2 to regime 3 and consequently will increase the posterior probability that the inputs to the fixed-contrast eye are outliers. This effectively suppresses the fixed-contrast eye and reduces its dominance durations. When the fixed-contrast eye sees a low-contrast image and the contrast of the variable-contrast eye is increased to be higher, the system transitions

from regime 1 to regime 2, thereby increasing the dominance duration of
the variable-contrast eye.

## 5 Discussion

Perceptual multistability is perhaps the most compelling evidence available
that the brain attempts to approximately represent a posterior distribution,
rather than just a point estimate, about the latent causes underlying sen-
sory inputs. The theory developed in this article, though incomplete in
many ways, is proof of concept that sampling-based posterior approxima-
tions on simple but realistic graphical models can reproduce a variety of
psychophysical phenomena, including the rich spatiotemporal dynamics
observed in binocular rivalry experiments. Importantly, these psychophys-
ical phenomena arise in our model from an approximately rational solution
to the fundamental problem of inferring the correct image interpretation
from noisy sensory inputs.

**5.1 Related Work.** The idea that perceptual multistability can be con-
strued in terms of sampling in a Bayesian model was first proposed by
Sundareswara and Schrater (Schrater & Sundareswara, 2007; Sundareswara
& Schrater, 2008), and our work follows theirs closely in several respects.
However, we believe our work offers several additional theoretical in-
sights. First, Sundareswara and Schrater's model requires sampling from
the known full posterior; thus, it is not clear whether their model reflects an
effective approximate inference algorithm when the posterior is not known.
In contrast, we show that this behavior need not require sampling from the
full posterior, but emerges from a simple, common, and effective sampling
method for approximate inference in graphical models, demonstrating that
multistable perception can be understood as the consequence of the engi-
neering principles responsible for recent progress in computer vision and
artificial intelligence (Geman & Geman, 1984; Koller & Friedman, 2009).
Second, our model offers an explanation for the memory decay on the lat-
est best sample postulated by Sundareswara and Schrater: this apparent
persistence and decay of sampled states is a natural consequence of the
principled coupling of adjacent eye and image nodes.

These principles highlight a more general point of divergence between
our model and other Bayesian models of multistable perception (Dayan,
1998; Hohwy, Roepstorff, & Friston, 2008). Other models postulate in-
trinsically stochastic neurophysiological processes such as noisy sensory
detection, adaptation, inhibition, and so on, to account for the stochastic
dynamics of multistability. In contrast, we show that these stochastic dy-
namics fall out naturally from a rational approach to approximate Bayesian
inference in graphical models. To be clear, we do not mean to suggest that
the details of stochastic neurophysiological processes are unimportant (see
section 5.2), but our goal here is to show how the stochastic dynamics

of perceptual multistability could arise from rational algorithms for approximate probabilistic inference, independent of—and at a more abstract level of analysis than—any specific mechanism for implementing inference algorithms in the brain. Relating our proposals at the algorithmic level to the details of neural processing is an important avenue for future work.

Successful dynamic neural models of binocular rivalry have argued that dominance arises from inhibitory connections between monocular neurons that suppress one eye or the other, while switching arises from adaptation of those inhibitory neurons, reducing suppression of one eye and allowing it to suppress the formerly dominant eye in turn (Wilson, 2003; Noest, van Ee, Nijs, & Van Wezel, 2007). Such models are supported by the observation that suppression and dominance are complementary (Alais, Cass, O'Shea, & Blake, 2010): gradual loss of visual sensitivity in one eye is accompanied by gradual gain in the other eye. Our model suggests a rationale for such behavior: when the inferred latent image is inconsistent with the sensory signals in one eye, the visual system must explain away those sensory signals by ascribing them to a noise process. Thus, while an eye is suppressed, signals to that eye are interpreted as arising from noise, resulting in a loss of sensitivity and suppression of input to that eye. Moreover, this outlier process is coupled between eyes because inferring an outlier in both eyes is less parsimonious than simply explaining away the sensory input in one eye; thus, it is sensible for the two eyes to alternate and for the dominance of one to co-occur with the suppression of the other. Which outlier process is preferred depends on local propagation dynamics, and shifts from one state to another are enabled by stochastic changes in the surrounding nodes as well as in the latent image nodes. Thus, adaptation-like dynamics can arise from simple network dynamics. Of course, these two explanations are not mutually exclusive, and adaptation might be a neural heuristic for approximating the appropriate propagation dynamics.

Another important explanatory principle employed by a number of recent models is a hierarchy of processing levels at which perceptual competition can occur (Wilson, 2003; Freeman, 2005). This principle is motivated by findings from psychophysical experiments showing that competition occurs both between eyes and between objects. For example, rapid swapping of flickering orthogonal monocular gratings between the eyes results in perceptual dominance durations that exceed the rate of swapping (Logothetis, Leopold, & Sheinberg, 1996), suggesting that rivalry is at the level of perceptual objects rather than the eye of origin. However, in experiments where monocular stimuli are swapped after the subject reports exclusive dominance in one eye, the suppressed perceptual object immediately becomes dominant and the dominant pattern becomes suppressed (Blake, Westendorf, & Overton, 1980), suggesting that rivalry is at the level of eyes rather than objects. These conflicting findings are somewhat resolved by brain imaging studies providing evidence for competition in both V1 and in extrastriate areas (see Tong, Meng, & Blake, 2006, for

a review): rivalry in the brain appears to occur at the level of both the monoocular sensory input and the binocular perceptual object. Our model also captures this notion of hierarchical competition: rivalry at the ocular level arises from the spatial coupling of the eye-specific noise process MRFs, thus encouraging one or another eye to be dominant. In turn, rivalry at the image level arises from the coupling of the latent image MRFs, which captures binocular extrastriate representations and encourages single coherent images to be dominant. Because the sampling dynamics operate over MRFs at both levels, the multistable stochastic alternations are driven by changes at the eye and pattern level and reflect both ocular and pattern rivalry.

**5.2 Neural Implementation.** Many neurally plausible models have been proposed that emulate binocular rivalry (Laing & Chow, 2002; Freeman, 2005; Moreno-Bote et al., 2007; Noest et al., 2007; Wilson, 2007). We have set out to address a slightly different question that does not challenge the validity of these models: What does such a neural architecture accomplish? Our work in this article can be seen as a rational statistical analysis that gives insight into the neural dynamics of these models why they work as they do. In this section, we bring the focus back to neural architecture, highlighting the similarity between our model and previously proposed architectures. Explicit links can be established between the rational dynamics of inference in an MRF and the dynamics in classic neural network models, including models of rivalry.

There is an intimate connection between the MRF model presented here and classical neural networks in computational neuroscience, such as the Hopfield net (Hopfield, 1982) and the Boltzmann machine (Ackley, Hinton, & Sejnowski, 1985). In fact, both the Hopfield net and the Boltzmann machine are examples of MRFs. The latent scene nodes in our model can be thought of as neuron-like units whose "firing" signals the presence of a particular image feature and whose "synaptic input" is the negative energy potential contributed by local neurons and sensory inputs (the retinal image nodes). This local connectivity structure is characteristic of primary visual cortex (Das & Gilbert, 1999). Another interesting aspect of this model is that the negative energy potential integrated by each neuron is a logarithmic function of probability; such logarithmic coding has been widely implicated in cortical systems (Gold & Shadlen, 2002). Furthermore, the stochastic nature of neural firing makes it well suited for implementing sample-based approximations (Hoyer & Hyvärinen, 2003).

Gibbs sampling operating over the MRF gives rise to what is known as Glauber dynamics (Glauber, 1963), a well-studied phenomenon in the neural networks literature (Amit, 1989). The hallmark of Glauber dynamics (as applied here) is spatiotemporal autocorrelation in neural responses (e.g., traveling waves) as the network diffuses stochastically toward low-energy stable states. Importantly, once a stable state is reached, the stochastic nature of the dynamics ensures that eventually the network will leave that state

and temporarily settle in a new one, thereby anticipating the dynamics of perceptual multistability.

To make these connections more explicit and show the link to neurally based models of rivalry, consider the simplified setting where $s_n \in \{x_n^L, x_n^R\}$, which is equivalent to assuming that $w_n \in \{0, 1\}$, and $\pi_n = 1, \forall n$. That is, $w_n = 1$ if the left eye is dominant at pixel $n$, and 0 otherwise. The Gibbs sampling updates presented in the appendix can be computed by a stochastic network of neurons. Let $y_n$ denote a binary neuron whose firing ($y_n = 1$) signals that the left eye is dominant in the latent image at pixel $n$. This neuron receives net synaptic input $a_n$ and emits a spike with probability

$$P(y_n = 1) = \frac{1}{1 + e^{-\tau a_n}}, \tag{5.1}$$

where

$$a_n = -\left(b_n - x_n^L\right)^2 - \left(b_n - x_n^R\right)^2 - \beta \sum_{j \in \mathcal{C}_n} (2w_j - 1). \tag{5.2}$$

The synaptic inputs consist of lateral connections from neighboring binocular neurons ($w_j$) and monocular retinal inputs ($x_n^L$ and $x_n^R$). To a first approximation, the net synaptic input reflects competition between eyes as well as between stimulus patterns (as in Wilson, 2003). In particular, the contrastive terms in the activation function can be interpreted as mutually inhibitory connections.

Another property embodied by these equations is the stochastic nature of spike generation that enables transitions between attractor states. The role of noise-induced transitions in multistability has been emphasized by several recent models (Moreno-Bote et al., 2007; van Ee, 2009). Our rational analysis elucidates the role that such transitions might play in allowing the visual system to sample statistically optimal percepts in a computationally tractable fashion.

It is straightforward to verify that sampling from equation 5.1 is equivalent to sampling from equation A.5 in the appendix. Thus, perceptual inference in our model can be mapped approximately onto a linear-nonlinear Poisson cascade (Schwartz, Pillow, Rust, & Simoncelli, 2006): (1) a linear integration of synaptic inputs, (2) a nonlinear transformation of the net input into a scalar firing rate, and (3) stochastic spike generation according to a Poisson process, approximated in discrete time by a Bernoulli process. The linear-nonlinear Poisson cascade is a canonical building block of neural models, particularly in early vision (Carandini et al., 2005); our model provides a functional interpretation of the cascade in terms of rational perceptual inference.

**5.3 Limitations and Extensions.** MRFs are just one of many kinds of probabilistic graphical model that have been successfully applied in recent computer vision and AI research. They are well suited to lower-level perceptual tasks but not to more causally structured higher-level problems of scene understanding; there, directed models such as Bayesian networks are more appropriate (Yuille & Kersten, 2006). Likewise, multistability occurs in other perceptual contexts besides binocular rivalry, such as depth reversals in the perception of three-dimensional objects (Blake & Logothetis, 2002). It would be interesting to see if phenomena of multistability occurring in higher-level perceptual contexts can also be explained in terms of MCMC inference on appropriately structured graphical models.

While our account emphasizes the algorithmic dynamics of sampling, an alternative perspective, explored by Bialek & DeWeese (1995), is that multistability arises from the dynamics of the sensory inputs themselves. This perspective holds that multistability is a normative adaptation to a nonstationary world. Although intuitively appealing given that natural sensory inputs really are nonstationary, a limitation of this persepctive is that it does not apply as congenially to situations where ambiguity is due to static causes, for example, occlusion of one eye.

A more fundamental weakness of our model pertains to its representation of uncertainty. MCMC methods represent uncertainty through the ensemble of samples generated over time, yet we have portrayed the perceptual system as essentially keeping track of only the most recent sample. One way to rehabilitate the normativity of the model's output would be to posit a downstream area (e.g., in extrastriate or inferior temporal cortex) that compiles the samples or estimates statistics in an online fashion. We leave development of such a theory to future work.

Although we have focused on perceptual multistability, other cognitive processes may also be amenable to a similar analysis. For example, it has been suggested that the acquisition of logical concepts could arise from MCMC-based diffusion through a space of grammars (Ullman, Goodman, & Tenenbaum, 2010). Anchoring and adjustment effects, in which people seem to incrementally adjust their numerical estimates away from an anchor (Epley & Gilovich, 2006), could arise from a similar process. Generally MCMC methods make inductive inference practical in many different kinds of hypothesis spaces, and their characteristic dynamics lead to testable predictions.

In conclusion, we have argued that MCMC algorithms provide a psychologically and biologically plausible explanation for perceptual multistability where traditional point estimation schemes fail. It is worth emphasizing that purely computational-level analyses will not suffice to explain this phenomenon; the explanatory power of our model derives from its algorithmic-level analysis. Although incomplete in many ways, it points toward the fruitfulness of bringing algorithmic ideas into rational statistical theories of cognition.

## Appendix: Model Details

**A.1 Image Model.** We formulate the posterior distribution over image and outlier variables using a Gibbs distribution:

$$P(\mathbf{s}, \pi | \mathbf{x}) \propto \exp\left\{-\tau \sum_{n=1}^{N} E_n^s - \tau \sum_i \left(E_n^{xi} + E_n^{\pi i}\right)\right\}, \tag{A.1}$$

where $i \in \{L, R\}$ indexes the eyes, $\tau$ is a global inverse temperature parameter (which we set to 1/250), and the energy functions are defined as follows:

$$E_n^{xi} = \frac{\pi_n^i \left(x_n^i - s_n\right)^2}{2\sigma_i^2}, \tag{A.2}$$

$$E_n^s = (b_n - s_n)^2 + \beta \sum_{j \in \mathcal{C}_n} (w_n - w_j)^2, \tag{A.3}$$

$$E_n^{\pi i} = \sum_{j \in \mathcal{C}_n} \gamma \left(\pi_n^i - \pi_j^i\right)^2 + \alpha \pi_j^i, \tag{A.4}$$

where $s_n = w_n x_n^L + (1 - w_n) x_n^R$ and $\mathcal{C}_n$ denotes the subset of pixel locations neighboring $n$. The parameters $\beta$ and $\gamma$ control the smoothness of the image and outlier MRFs, respectively; we set both of these to 10 in our simulations. The parameter $\alpha$ is a sparsity parameter that penalizes outliers; we set this parameter to 500. The parameter $b_n$ is a bias term for the $n$th pixel, encoding prior knowledge about the latent surface at each location, which is set to 0 except where noted otherwise. The variance parameter $\sigma_i^2$ was set to 0.005 except where noted otherwise.

**A.2 Conditional Distributions.** In the Gibbs sampler, nodes are updated asynchronously by drawing their values from the following conditional distributions:

$$P(w_n | \mathbf{x}, \pi, \mathbf{w}_{/n}) \propto \exp\left\{-\tau E_n^s - \tau \sum_i E_n^{xi}\right\}, \tag{A.5}$$

$$P\left(\pi_n^i | \mathbf{x}, \pi_{/n}^i, \mathbf{w}_n\right) \propto \exp\left\{-\tau E_n^{xi} - \tau E_n^{\pi i}\right\}, \tag{A.6}$$

where $\mathbf{s}_{/n}$ denotes all the nodes except the $n$th node (and likewise for $\pi$). Note that because $s_n$ is a deterministic function of $w_n$ and $\mathbf{x}_n$, sampling from $P(w_n | \mathbf{x}, \pi, \mathbf{w}_{/n})$ is equivalent to sampling from $P(s_n | \mathbf{x}, \pi, \mathbf{s}_{/n})$.
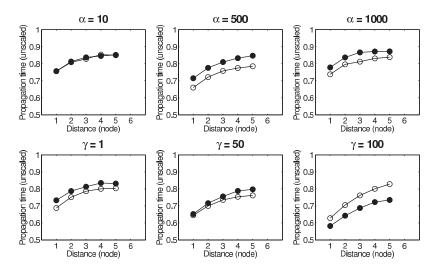
Figure 7: Sensitivity to sparsity level and outlier smoothness. Simulated traveling waves (cf. Figure 3) at different settings of $\alpha$ (top) and $\gamma$ (bottom).

**A.3 Parameter Selection and Robustness.** The parameters of our model were chosen heuristically so as to approximately match the statistics of the empirically observed dominance duration distribution (see Figure 2). These parameters were then used in all other simulations, except where explicitly manipulated. In principle, one could use stochastic approximation techniques (Swersky, Chen, Marlin, & de Freitas, 2010) to optimize the parameters, but the computation time required to obtain a sufficiently large set of samples makes that approach prohibitively expensive.

To illustrate robustness to variations in parameter settings, we recalculated the predicted traveling waves under different values of $\alpha$ (the sparsity parameter) and $\gamma$ (outlier smoothness parameter).[1] Figure 7 shows the results of these simulations. While the quantitative speed of propagation changes slightly as a function of $\alpha$ and $\gamma$, the qualitative pattern shown in Figure 3 remains unchanged; the only exception is for $\gamma = 100$, where the difference between concentric and radial propagation times reverses (this is a somewhat degenerate case where the coupling of the outlier process is an order of magnitude greater than the cost of inducing an outlier, $\alpha$). Thus, model predictions appear to be relatively insensitive to variations in the sparsity level and outlier smoothness.

---

[1]Note that we have already explicitly manipulated $\beta$ in explaining the difference between radial and concentric propagation times.

## Acknowledgments

## References

Ackley, D., Hinton, G., & Sejnowski, T. (1985). A learning algorithm for Boltzmann machines. *Cognitive Science*, *9*(1), 147–169.

Alais, D., & Blake, R. (1999). Grouping visual features during binocular rivalry. *Vision Research*, *39*(26), 4341–4353.

Alais, D., Cass, J., O'Shea, R., & Blake, R. (2010). Visual sensitivity underlying changes in visual consciousness. *Current Biology*, *20*, 1362–1367.

Amit, D. (1989). *Modeling brain function: The world of attractor neural networks*. Cambridge: Cambridge University Press.

Bialek, W., & DeWeese, M. (1995). Random switching and optimal processing in the perception of ambiguous signals. *Physical Review Letters*, *74*(15), 3077–3080.

Blake, R. (2001). A primer on binocular rivalry, including current controversies. *Brain and Mind*, *2*(1), 5–38.

Blake, R., & Logothetis, N. (2002). Visual competition. *Nature Reviews Neuroscience*, *3*(1), 13–21.

Blake, R., O'Shea, R. P., & Mueller, T. J. (1992). Spatial zones of binocular rivalry in central and peripheral vision. *Visual Neuroscience*, *8*, 469–478.

Blake, R., Westendorf, D., & Overton, R. (1980). What is suppressed during binocular rivalry? *Perception*, *9*(2), 223–231.

Brainard, D., & Freeman, W. (1997). Bayesian color constancy. *Journal of the Optical Society of America A*, *14*(7), 1393–1411.

Brascamp, J., van Ee, R., Noest, A., Jacobs, R., & van den Berg, A. (2006). The time course of binocular rivalry reveals a fundamental role of noise. *Journal of Vision*, *6*(11), 8.

Burke, D., Alais, D., & Wenderoth, P. (1999). Determinants of fusion of dichoptically presented orthogonal gratings. *Perception*, *28*, 73–88.

Carandini, M., Demb, J., Mante, V., Tolhurst, D., Dan, Y., Olshausen, B., et al. (2005). Do we know what the early visual system does? *Journal of Neuroscience, 25*(46), 10577–10587.

Das, A., & Gilbert, C. (1999). Topography of contextual modulations mediated by short-range interactions in primary visual cortex. *Nature*, *399*(6737), 655–661.

Daw, N., & Courville, A. (2007). The pigeon as particle filter. In J. C. Platt, D. Koller, Y. Singer, & S. Roweis (Eds.), *Advances in neural information processing systems, 20* (pp. 369–376). Cambridge, MA: MIT.

Dayan, P. (1998). A hierarchical model of binocular rivalry. *Neural Computation*, *10*(5), 1119–1135.

Epley, N., & Gilovich, T. (2006). The anchoring-and-adjustment heuristic. *Psychological Science*, *17*(4), 311–318.

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.

Fox, R., & Check, R. (1968). Detection of motion during binocular rivalry suppression. *Journal of Experimental Psychology*, *78*(3), 388–395.

Freeman, A. (2005). Multistage model for binocular rivalry. *Journal of Neurophysiology*, *94*(6), 4412–4420.

Fukuda, H. (1981). Magnitude of suppression of binocular rivalry within the invisible pattern. *Perceptual and Motor Skills*, *53*(2), 371–375.

Geman, S., & Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions of Pattern Analysis and Machine Intelligence*, *6*, 721–741.

Glauber, R. (1963). Time-dependent statistics of the Ising model. *Journal of Mathematical Physics*, *4*, 294–307.

Gold, J., & Shadlen, M. (2002). Banburismus and the brain: Decoding the relationship between sensory stimuli, decisions, and reward. *Neuron*, *36*(2), 299–308.

Griffiths, T., & Tenenbaum, J. (2006). Optimal predictions in everyday cognition. *Psychological Science*, *17*(9), 767–773.

Grinband, J., Hirsch, J., & Ferrera, V. (2006). A neural representation of categorization uncertainty in the human brain. *Neuron*, *49*(5), 757–763.

Hohwy, J., Roepstorff, A., & Friston, K. (2008). Predictive coding explains binocular rivalry: An epistemological review. *Cognition*, *108*(3), 687–701.

Hollins, M. (1980). The effect of contrast on the completeness of binocular rivalry suppression. *Attention, Perception, and Psychophysics*, *27*(6), 550–556.

Hopfield, J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences of the United States of America*, *79*(8), 2254–2258.

Hoyer, P., & Hyvärinen, A. (2003). Interpreting neural response variability as Monte Carlo sampling of the posterior. In S. Becker, S. Thrün, & K. Obermayer (Eds.), *Advances in neural information processing systems, 15* (pp. 277–284). Cambridge, MA: MIT Press.

Kang, M. (2009). Size matters: A study of binocular rivalry dynamics. *Journal of Vision*, *9*(1), 1–11.

Kiani, R., & Shadlen, M. (2009). Representation of confidence associated with a decision by neurons in the parietal cortex. *Science*, *324*(5928), 759–764.

Knapen, T., Kanai, R., Brascamp, J., van Boxtel, J., & van Ee, R. (2007). Distance in feature space determines exclusivity in visual rivalry. *Vision Research*, *47*(26), 3269–3275.

Knill, D., & Pouget, A. (2004). The Bayesian brain: The role of uncertainty in neural coding and computation. *Trends in Neurosciences*, *27*(12), 712–719.

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

Koller, D., & Friedman, N. (2009). *Probabilistic graphical models: Principles and techniques*. Cambridge, MA: MIT Press.

Kording, K., & Wolpert, D. (2004). Bayesian integration in sensorimotor learning. *Nature*, *427*(6971), 244–247.

Laing, C., & Chow, C. (2002). A spiking neuron model for binocular rivalry. *Journal of Computational Neuroscience*, *12*(1), 39–53.

Lankheet, M. (2006). Unraveling adaptation and mutual inhibition in perceptual rivalry. *Journal of Vision*, *6*(4), 304–310.

Lee, S., Blake, R., & Heeger, D. (2005). Traveling waves of activity in primary visual cortex during binocular rivalry. *Nature Neuroscience*, *8*(1), 22–23.

Lee, T., & Mumford, D. (2003). Hierarchical Bayesian inference in the visual cortex. *Journal of the Optical Society of America A*, *20*(7), 1434–1448.

Leopold, D., & Logothetis, N. (1996). Activity changes in early visual cortex reflect monkeys' percepts during binocular rivalry. *Nature*, *379*(6565), 549–553.

Levelt, W. (1965). *On binocular rivalry.* Soesterberg, Netherlands: Institute for Perception Rvo-Tno.

Levy, R., Reali, F., & Griffiths, T. (2009). Modeling the effects of memory on human online sentence processing with particle filters. In D. Koller, D. Schuurmans, Y. Bengio, & L. Bottou (Eds.), *Advances in neural information processing systems, 21* (pp. 937–944). Cambridge, MA: MIT Press.

Liu, L., Tyler, C., & Schor, C. (1992). Failure of rivalry at low contrast: Evidence of a suprathreshold binocular summation process. *Vision Research*, *32*(8), 1471–1479.

Logothetis, N., Leopold, D., & Sheinberg, D. (1996). What is rivalling during binocular rivalry? *Nature*, *380*(6575), 621–624.

Mamassian, P., & Goutcher, R. (2005). Temporal dynamics in bistable perception. *Journal of Vision*, *5*(4), 361–375.

Moreno-Bote, R., Rinzel, J., & Rubin, N. (2007). Noise-induced alternations in an attractor network model of perceptual bistability. *Journal of Neurophysiology*, *98*(3), 1125–1139.

Mueller, T., & Blake, R. (1989). A fresh look at the temporal dynamics of binocular rivalry. *Biological Cybernetics*, *61*(3), 223–232.

Noest, A., van Ee, R., Nijs, M., & Van Wezel, R. (2007). Percept-choice sequences driven by interrupted ambiguous stimuli: A low-level neural model. *Journal of Vision*, *7*(8), 1–14.

O'Shea, R. (1987). Chronometric analysis supports fusion rather than suppression theory of binocular vision. *Vision Research*, *27*(5), 781–791.

O'Shea, R., Sims, A., & Govan, D. (1997). The effect of spatial frequency and field size on the spread of exclusive visibility in binocular rivalry. *Vision Research*, *37*(2), 175–183.

Polonsky, A., Blake, R., Braun, J., & Heeger, D. (2000). Neuronal activity in human primary visual cortex correlates with perception during binocular rivalry. *Nature Neuroscience*, *3*(11), 1153–1159.

Robert, C., & Casella, G. (2004). *Monte Carlo statistical methods*. New York: Springer.

Sanborn, A., Griffiths, T., & Navarro, D. (2006). Rational approximations to rational models: Alternative algorithms for category living. *Psychological Review*, *117*, 1144–1167.

Schrater, P., & Kersten, D. (2000). How optimal depth cue integration depends on the task. *International Journal of Computer Vision*, *40*(1), 71–89.

Schrater, P., & Sundareswara, R. (2007). Theory and dynamics of perceptual bistability. In B. Schölkopf, J. Platt, & T. Hoffman (Eds.), *Advances in neural information processing systems*, *19* (pp. 1217–1224). Cambridge, MA: MIT Press.

Schwartz, O., Pillow, J., Rust, N., & Simoncelli, E. (2006). Spike-triggered neural characterization. *Journal of Vision*, *6*(4), 484–507.

Shi, L., & Griffiths, T. (2009). Neural implementation of hierarchical Bayesian inference by importance sampling. In Y. Bengio, D. Schuurmans, J. Lafferty, C. K. I. Williams, & A. Culotta (Eds.), *Advances in neural information processing systems, 22* (pp. 1669–1677). Cambridge, MA: MIT Press.

Shimojo, S., & Nakayama, K. (1990). Real world occlusion constraints and binocular rivalry. *Vision Research*, *30*(1), 69–80.

Sundareswara, R., & Schrater, P. (2008). Perceptual multistability predicted by search model for Bayesian decisions. *Journal of Vision*, *8*(5), 1–19.

Swersky, K., Chen, B., Marlin, B., & de Freitas, N. (2010). A tutorial on stochastic approximation algorithms for training restricted Boltzmann machines and deep belief nets. In *Information Theory and Applications Workshop, 2010* (pp. 1–10). Piscataway, NJ: IEEE.

Tong, F., Meng, M., & Blake, R. (2006). Neural bases of binocular rivalry. *Trends in Cognitive Sciences*, *10*(11), 502–511.

Ullman, T., Goodman, N., & Tenenbaum, J. (2010). Theory acquisition as stochastic search. In *Proceedings of the 32nd Annual Cognitive Science Society*. Cognitive Science Society.

van Ee, R. (2009). Stochastic variations in sensory awareness are driven by noisy neuronal adaptation: Evidence from serial correlations in perceptual bistability. *JOSA A*, *26*(12), 2612–2622.

Vul, E., Goodman, N., Griffiths, T., & Tenenbaum, J. (2009). One and done? Optimal decisions from very few samples. In *Proceedings of the 31st Annual Meeting of the Cognitive Science Society*. Cognitive Science Society.

Wales, R., & Fox, R. (1970). Increment detection thresholds during binocular rivalry suppression. *Perception and Psychophysics*, *8*(2), 90–94.

Weiss, Y., Simoncelli, E., & Adelson, E. (2002). Motion illusions as optimal percepts. *Nature Neuroscience*, *5*(6), 598–604.

Wilson, H. (2003). Computational evidence for a rivalry hierarchy in vision. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(24), 14499–14503.

Wilson, H. (2007). Minimal physiological conditions for binocular rivalry and rivalry memory. *Vision Research*, *47*(21), 2741–2750.

Wilson, H., Blake, R., & Lee, S. (2001). Dynamics of travelling waves in visual perception. *Nature*, *412*(6850), 907–910.

Yuille, A., & Kersten, D. (2006). Vision as Bayesian inference: Analysis by synthesis? *Trends in Cognitive Sciences*, *10*(7), 301–308.

---