Bayesian sampling in visual perception

Rubén Moreno-Bote[a,b,1], David C. Knill[b,c], and Alexandre Pouget[b]

[a]Foundation Sant Joan de Déu, Parc Sanitari Sant Joan de Déu, Esplugues de Llobregat, 08950 Barcelona, Spain; and [b]Department of Brain and Cognitive Sciences and [c]Center for Visual Science, University of Rochester, Rochester, NY, 14627

Edited by Wilson S. Geisler, University of Texas at Austin, Austin, TX, and approved May 31, 2011 (received for review January 27, 2011)

It is well-established that some aspects of perception and action can be understood as probabilistic inferences over underlying probability distributions. In some situations, it would be advantageous for the nervous system to sample interpretations from a probability distribution rather than commit to a particular interpretation. In this study, we asked whether visual percepts correspond to samples from the probability distribution over image interpretations, a form of sampling that we refer to as Bayesian sampling. To test this idea, we manipulated pairs of sensory cues in a bistable display consisting of two superimposed moving drifting gratings, and we asked subjects to report their perceived changes in depth ordering. We report that the fractions of dominance of each percept follow the multiplicative rule predicted by Bayesian sampling. Furthermore, we show that attractor neural networks can sample probability distributions if input currents add linearly and encode probability distributions with probabilistic population codes.

Bayesian inference | neuronal network | neuronal noise | perceptual bistability

NEUROSCIENCE

PSYCHOLOGICAL AND COGNITIVE SCIENCES

There is mounting evidence that neural circuits can implement probabilistic inferences over sensory, cognitive, or motor variables. In some cases, humans can perform these inferences optimally, as in multi-cue or multisensory integration (1–8). For complex tasks, such as object recognition, action perception, and object tracking, the computations required for optimal inference are intractable, which implies that humans must use approximate inferences (9–11). One approximate scheme that is particularly appealing from a biological point of view is sampling. Consider as an example the problem of object recognition. The goal of the inference in this case would be to compute the probability over object identities given the image. Although this probability distribution may be difficult to compute explicitly, one can often design algorithms to generate samples from the distribution, allowing one to perform approximate inference (12, 13). Some human cognitive choice behaviors suggest that the nervous system implements sampling. However, whether the same is true for low-level perceptual processing is currently unknown.

Stimuli that lead to bistable percepts (14–18), like the Necker cube, provide a tractable experimental preparation for testing the sampling hypothesis. With such stimuli, perception alternates stochastically between two possible interpretations, a behavior consistent with sampling as suggested by several works (16, 19, 20). However, the key question is what probability distribution is being sampled. If the brain uses sampling for Bayesian inference, neural circuits should sample from an internal probability distribution on possible stimulus interpretations that are conditioned on the available sensory data, the so-called posterior distribution. This distribution places important constraints on the distributions of perceptual states for bistable stimuli.

To test this idea, we used stimuli composed of two drifting gratings whose depth ordering is ambiguous (21). We then manipulated two depth cues to vary the fractions of dominance of the percepts. Our central prediction is that the fractions of dominance of each percept should behave as probabilities if they are the result of a sampling process of a posterior distribution over image interpretations. We will refer to this form of sampling as Bayesian sampling. First, we show that subjects' fractions of dominance in different cue conditions follow the same multiplicative rule as probabilities in the Bayesian calculus, suggesting that bistable perception is indeed a form of Bayesian sampling. Second, we describe possible neural implementations of a Bayesian sampling process using attractor networks, and we discuss the link with probabilistic population codes (22).

Results

Multiplicative Rule for Combining Empirical Fractions of Dominance. We asked subjects to report their spontaneous alternations in perceived depth ordering of two superimposed moving gratings over a 1-min period and measured the fraction of dominance time for each percept (*Methods* and Fig. 1*A*). In the first experiment, the two drifting gratings, α and β, were parameterized by their wavelength and speed. One of the wavelengths was always set to a fixed value λ^*, and one of the speeds was set to a fixed value v^*. The remaining wavelength and speed parameters, λ and v, respectively, determined the difference in wavelength and speed between gratings α and β, denoted $\Delta\lambda$ and Δv, and hence, the information for choosing grating α as the one behind. We refer to these differences as the cues to depth ordering, and we refer to the condition where the two differences are zero as the neutral cue condition ($\Delta\lambda = 0$ and $\Delta v = 0$). These cues have been shown to have a strong effect on the depth ordering of the gratings because of their relationship with the natural statistics of wavelength and speed of distant objects (21). In the second experiment, we manipulated wavelength and disparity, d, of the gratings. In this case, the label v should be interchanged with the label d.

According to the Bayesian sampling hypothesis, the empirical fractions of dominance arise from a process that samples the posterior distribution on possible scene interpretations given the sensory input. As we show in *SI Methods*, when two conditionally independent cues are available (i.e., the values of the cues are independent when conditioned on true depth), an optimal system should sample from a probability distribution given by the normalized product of the probability distributions derived by varying each cue in isolation while keeping the other cue neutral. Our hypothesis implies that the empirical fractions should behave as probabilities, and therefore, they should follow the multiplicative rule (Eq. **1**)

$$f_{\lambda v} = \frac{f_\lambda f_v}{f_\lambda f_v + (1-f_\lambda)(1-f_v)}, \qquad [1]$$

where $f_{\lambda v}$ is the fraction of time that subjects report percept A (grating α moving behind grating β) when the cues are set to $\Delta\lambda$ and Δv, f_λ is the fraction of dominance of percept A when the speed cue is neutral ($\Delta v = 0$) while the wavelength cue has value $\Delta\lambda$, and f_v is the dominance fraction when the wavelength cue is neutral ($\Delta\lambda = 0$) while the speed cue has value Δv. This relation holds whether subjects are sampling from posterior distributions

Author contributions: R.M.-B., D.C.K., and A.P. designed research; R.M.-B., D.C.K., and A.P. performed research; R.M.-B. analyzed data; and R.M.-B., D.C.K., and A.P. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence should be addressed. E-mail: rmoreno@bcs.rochester.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1101430108/-/DCSupplemental.

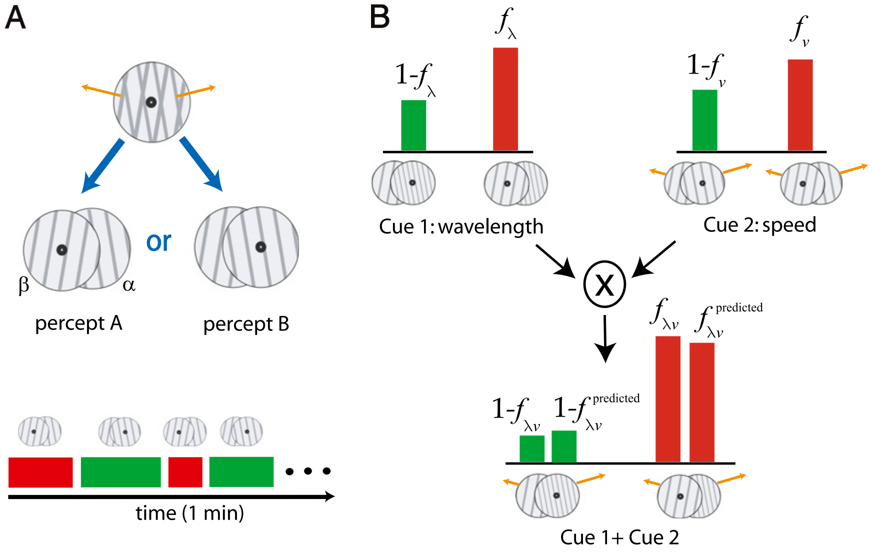

Fig. 1. Cue combination in a perceptually bistable stimulus. (*A*) The visual stimulus consisted of two superimposed drifting gratings moving in different directions. The perceived depth ordering of the gratings is bistable. We measured the fraction of dominance of each percept by asking subjects to report the perceived depth ordering of the gratings during trials of 1-min duration (hypothetical trial shown). (*B*) Cue combination. (*Upper Left*) Fractions of dominance for each depth ordering when wavelength is nonneutral (its value differs between the two gratings), whereas speed is neutral (its value is identical across gratings). *Upper Right* is the same as *Upper Left*, but when speed is nonneutral, the wavelength is neutral. (*Lower*) Fraction of dominance when both speed and wavelength are nonneutral. Bayesian sampling predicts that the fraction of dominance when both cues are nonneutral is equal to the normalized product of the fractions of dominance when only one cue is nonneutral (Eq. **1**). In the example illustrated here, both cues were congruent.

on depth or posterior distributions raised to an arbitrary power n (*SI Methods*). The multiplicative rule provides an empirical consistency constraint for Bayesian sampling. Note that this rule does not specify how the samples are extracted over time [i.e., it works whether the samples are independent over time (23, 24) or correlated]. As discussed later, bistable perception is only consistent with a sampling mechanism that generates correlated samples (i.e., the percept tends to remains the same over hundreds of milliseconds).

Observed vs. Predicted Fractions of Dominance. The multiplicative rule was tested in two experiments. In the first experiment, the wavelength and speed differences between the two gratings, $\Delta\lambda$ and Δv, were changed from trial to trial congruently [C condition (i.e., both cues favoring the same depth ordering); example in Fig. 1*B*] or incongruently [IC condition (i.e., the cues favored different depth orderings)]. This change was achieved by decreasing the wavelength and increasing the speed of grating α in the C condition, while decreasing the wavelength of grating α and increasing the speed of grating β in the IC condition. In the second experiment, the wavelength and stereo disparity (instead of speed) of the gratings were manipulated in the C and IC conditions as in the previous experiment.

As shown in Figs. 2 and 3, wavelength, speed, and disparity differences in the gratings have a strong impact on the fractions of dominance of the gratings' depth ordering (21). The fraction of dominance of percept *A* (grating α is behind grating β) increases as the wavelength difference between gratings α and β ($\Delta\lambda = \lambda_\alpha - \lambda_\beta$) decreases. The fraction increases as the speed difference between gratings α and β ($\Delta v = v_\alpha - v_\beta$) increases in the C condition (Fig. 2*A*). Conversely, the fraction decreases as the difference (in speed or wavelength) between the gratings decreases in the IC condition (Fig. 2*B*). In the second experiment, the fraction of dominance of percept *A* increases as the disparity difference between gratings α and β ($\Delta d = d_\alpha - d_\beta$) increases in the C condition (Fig. 3*A*). Again, the reverse pattern is observed in the IC condition (Fig. 3*B*). In the two experiments, when the two cues are set to their neutral values, the fractions (Figs. 2 and 3, black open circles) are not significantly different from one-half [two-tailed t test; experiment 1: $p = 0.39$ (C), $p = 0.06$ (IC) and experiment 2: $p = 0.31$ (C), $p = 0.051$ (IC)].

The experimental results were compared with the theoretical predictions from the multiplicative rule (Eq. **1**) (Figs. 2 *A* and *B* and 3 *A* and *B*). The predictions when the two cues are nonneutral (Figs. 2 *A* and *B* and 3 *A* and *B*, filled blue circles) were computed using the experimental data of the single nonneutral cue cases only (Figs. 2 *A* and *B* and 3 *A* and *B*, open red circles). The case in which wavelength is the only nonneutral cue corresponds to the lower line of open circles in Figs. 2*A* and 3*A* and the upper line in Figs. 2*B* and 3*B* in both experiments. The cases in which speed (or disparity) is the only nonneutral cue correspond to the vertical line of open circles in the wavelength and speed (or disparity) experiment in Fig. 2*B* (Fig. 3*B* respectively). The match between the observed data points (filled red circles) and predictions is tight, even though the multiplicative rule is parameter-free and cannot be adjusted to match the experimental results (note that, for the sake of clarity, the blue dots have been slightly displaced to the right). The data in Figs. 2 *A* and *B* and 3 *A* and *B* were replotted in Figs. 2*C* and 3*C* to show the predicted fraction of dominance from the multiplicative model vs. the observed fraction when the two cues were nonneutral with the C (Figs. 2*C* and 3*C*, light blue dots) and IC (Figs. 2*C* and 3*C*, dark blue) conditions combined. The strong alignment of the data points along the unity line confirms that the multiplicative rule provides a tight fit to the data. Individual subjects also followed the multiplicative rule (*SI Methods* and Fig. S1).

We also tested alternative models to the multiplicative rule. In the first model, we assumed that integration between the cues does not take place—a strongest cue take all model. In this model, performance is driven by the cue with the lowest uncertainty: The fraction of dominance when both cues are varied together is set to that of the cue whose fraction when the cues are manipulated alone has the largest absolute value difference with respect to one-half (*SI Methods*). As shown in Figs. 2*D* and 3*D* (brown dots) this model fails to capture our experimental results. In the second model, we generated predictions from a realistic neuronal network (see *Results*, *Sampling with Realistic Neural Circuits*). When the input neurons to the network fired nonlinearly in response to the stimuli (25), the predictions of the model, which fit the single nonneutral cue conditions, substantially differed from the experimental data in the four nonneutral cues conditions (NL net) (Figs. 2*D* and 3*D*, orange dots). When the input neurons fired linearly (26), the predictions were identical to the multiplicative rule (L net) (Figs. 2*D* and 3*D*, blue dots). This result shows that the mere fact that a network can oscillate stochastically between two percepts in a way suggestive of sampling does not guarantee that it will also follow the multiplicative rule. Whether it does depends critically on how the inputs are combined, a point that we discuss more thoroughly below.

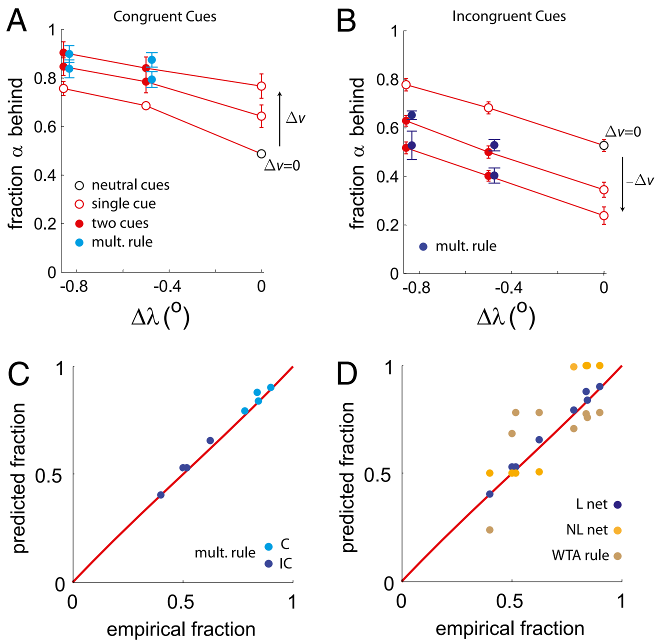

Fig. 2. Experimental and predicted fractions of dominance in the wavelength and speed cue combination experiment. (*A*) Fraction of dominance of percept *A* (i.e., grating α is behind grating β) as a function of the wavelength difference between gratings α and β ($\Delta\lambda = \lambda_\alpha - \lambda_\beta$) for three different speed differences ($\Delta v = v_\alpha - v_\beta$) in the congruent condition (both cues favored the same depth ordering). Data are averaged across subjects, and the error bars correspond to SEM across subjects. Experimental observations (red and black) and predictions from the multiplicative rule (blue circles) (Eq. **1**) are shown. The predictions from the multiplicative rule were computed using the experimental data from the conditions in which only one cue was nonneutral (open circles). The black open circles correspond to the fractions measured when the two cues were neutral. The predictions are displaced slightly right in relation to the experimental data (filled red circles) to allow better visual comparison. (*B*) Same as in *A* but for the incongruent condition (the cues favored opposite percepts). (*C*) Predicted fractions of dominance for the multiplicative rule combining the data from the congruent (C; light blue) and incongruent (IC; dark blue) conditions from *A* and *B* as a function of the empirical fractions. (*D*) Same as in *C* but for the strongest cue take all rule (brown) and a rate-based model with nonlinear (orange) and linear (blue) input neurons.

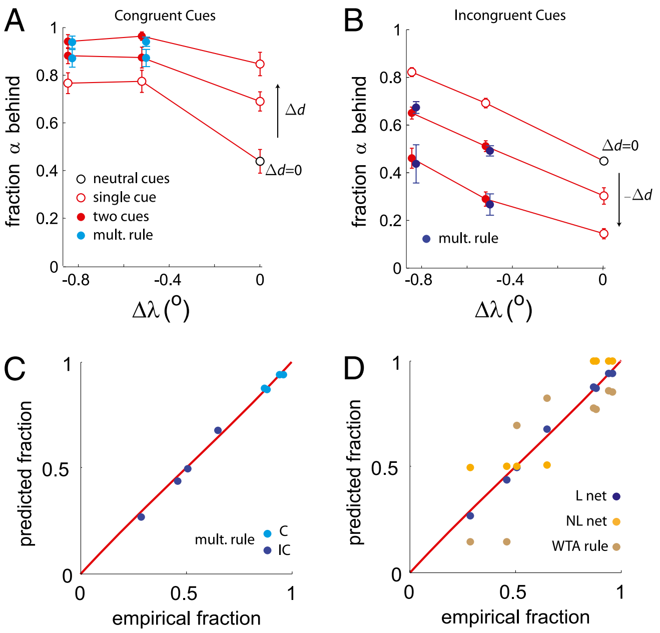

Fig. 3. Experimental and predicted fractions of dominance in the wavelength and disparity cue combination experiment. *A–D* are the same as in Fig. 2 but with speed replaced by disparity.

Diffusion in an Energy Model. Our finding that bistable perception behaves like a Bayesian sampling process raises the issue as to how neurons could implement such a process. We first show that implementing the multiplicative rule is surprisingly straightforward with energy models. In *Results*, *Sampling with Realistic Neural Circuits*, we will present a neural instantiation of this conceptual framework. We model the dynamics of two neural populations, *A* and *B*, whose states are described by their firing rates r_A and r_B, respectively (Fig. 4*A*). The reduced dynamics tracks the difference between the firing rates, $r = r_A - r_B$, where $r > 0$ corresponds to percept *A*. This variable obeys (Eq. **2**)

$$\tau \frac{d}{dt} r = -4r(r^2-1) + g(I_\lambda, I_v) + n(t), \qquad [2]$$

where $g(I_\lambda, I_v)$ is a bias provided by the inputs and $n(t)$ is a filtered white noise with variance σ^2 (27) (*SI Methods*). The first term on the right-hand side ensures that the activity difference, r, hovers around the centers of the two energy wells (Fig. 4*B*). The bias term measures the combined strength of the cues, which is a function of the individual strengths I_λ and I_v favoring percept *A* from the wavelength and speed cues, respectively. The function $g(I_\lambda, I_v)$ is chosen such that it is zero when the two cues are neutral (zero currents) and positive when the two cues favor percept *A* (the two currents are positive). The dynamics of Eq. **2** can be viewed as a noisy descent over the energy landscape $E(r) = r^2(r^2-2) - g(I_\lambda, I_v)r$, which is symmetrical (Fig. 4*B*, black line) when the two cues are neutral and negatively tilted (Fig. 4*B*, gray line) when the cues favor percept *A*. The resulting dynamics effectively draws samples from an underlying probability distribution that depends on the input currents (a process known as Langevin Monte Carlo sampling) (28).

To model the experimental data that we have described, we need a form of sampling that obeys the multiplicative rule. Whether the network obeys the rule or not depends critically on the function $g(I_\lambda, I_v)$. We consider here the family of functions described by $g(I_\lambda, I_v) = I_\lambda + I_v + \varepsilon(I_\lambda^2 I_v + I_v^2 I_\lambda)$, where ε measures the strength of the nonlinearity. Similar nonlinear functional dependences on the input currents naturally arise in neuronal networks with nonlinear activation functions (*Results*, *Sampling with Realistic Neural Circuits*).

For a value of ε different from zero, the dynamical system does not follow the multiplicative rule (Fig. 4*D*). In contrast, if we set ε to zero, such that $g(I_\lambda, I_v) = I_\lambda + I_v$, the system now obeys the multiplicative rule (Fig. 4*E*). This result can be derived analytically by computing the mean dominance duration of each percept, which corresponds to the mean escape time from one of the energy wells (*SI Methods*). We can then show that the fraction of dominance of population *A* for ε equal to zero is a sigmoid function of the sum of the inputs (Eq. **3**)

$$f_{\lambda v} = f(s = A \mid I_\lambda, I_v) = \frac{1}{1 + e^{-2(I_\lambda + I_v)/\sigma_{eff}^2}} \propto e^{(I_\lambda + I_v)/\sigma_{eff}^2}, \qquad [3]$$

where σ_{eff}^2 is the effective noise in the system and is proportional to σ^2. Note that when only one cue is nonneutral, $f_i \propto e^{I_i/\sigma_{eff}^2}$ ($i = \lambda$, v), and when both cues are nonneutral, $f_{\lambda v} \propto e^{(I_\lambda + I_v)/\sigma_{eff}^2}$. Therefore, the fractions are related through $f_{\lambda v} \propto f_\lambda \times f_v$, and after normalization, they follow the multiplicative rule (Eq. **1**). Fig. 4*F* shows that Eq. **3** is indeed satisfied by the diffusion model, because the fraction of dominance of percept *A* obtained from numerical simulations as a function of the total input current (Fig. 4*F*, blue line) is a sigmoid function (Fig. 4*F*, red line). This

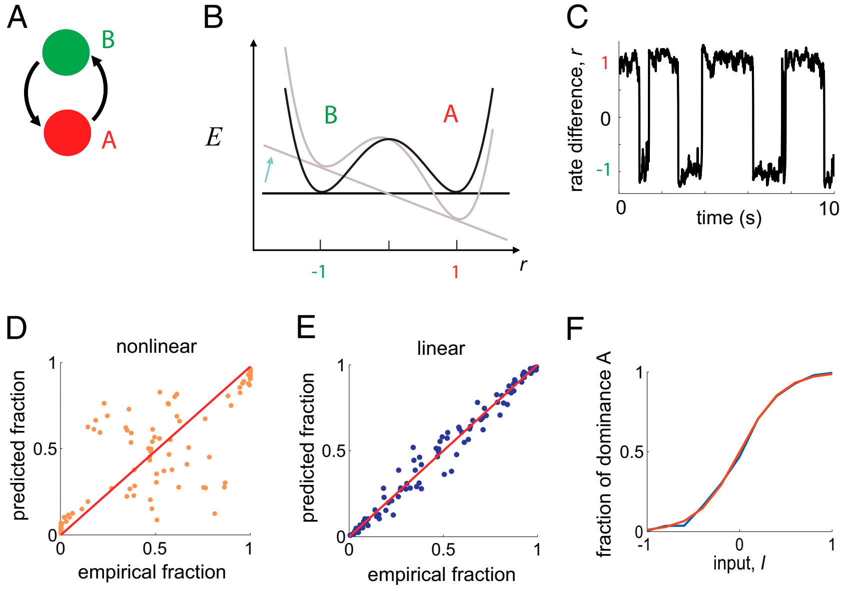

Fig. 4. Simplified network model for Bayesian sampling. (*A*) Schematic of the neural network. (*B*) Energy as a function of the difference between the firing rates of the two populations ($r = r_A - r_B$). When the state of the system lies close to the right or left minimum (r is close to 1 or −1), percept *A* or *B* dominates, respectively. Alternations in dominance happen because noise can kick the system from one minimum to the other minimum. When the two cues are neutral (black line), the two percepts dominate for equal amounts of time (i.e., $f = 0.5$). When the cues favor percept *A*, the energy landscape is tilted to the right (gray line), and $f > 0.5$. (*C*) Population rate difference r as a function of time. Stochastic switches occur between the two states of the system. (*D* and *E*) Fractions of dominance predicted by the multiplicative rule vs. observed fractions of dominance generated by the model (*D*, orange dots and *E*, blue dots) with nonlinear (*D*) and linear (*E*) inputs ($\varepsilon = 5$ and $\varepsilon = 0$, respectively). The model's performance lies close to the unit slope line (red) only when the inputs are combined linearly (*E*). (*F*) Fraction of dominance of state *A* ($r > 0$) as a function of the total input. The curve (blue) is well-fitted by a sigmoid function (red).

analytical approach can also be used to reveal why the system with a nonlinear function does not follow the multiplicative rule. Because in this case, $f_{\lambda\nu} \propto e^{g(I_\lambda, I_\nu)/\sigma^2_{eff}}$, the product of the fractions when only one cue is nonneutral is not equal to the fraction when the two cues are nonneutral.

Sampling with Realistic Neural Circuits. The main features of the energy model can be implemented in a neural network with attractor dynamics. We consider a recurrent neural network with two competing populations (Fig. 5*A*) encoding the two percepts *A* and *B*, whose states are described by their population averaged firing rates r_A and r_B, as suggested by neural data (29). An additional relay neuronal population fires in response to the cues and provides inputs to the competing populations *A* and *B* with positive (direct connections) and negative (through an inhibitory population) signs, respectively. The firing of the relay population is a function of the sum of the cue strengths, $I_\lambda + I_\nu$. We consider linear and nonlinear activation functions (*SI Methods*) close to those functions found in primary visual cortex (25, 26). We also added a slow adaptation process (30–33).

The network stochastically alternates between percepts with gamma-like distributions of dominance durations, which captures several aspects of the experimental distributions (Fig. 5*B*) (14, 17, 34–36). The distributions generated by the network are not significantly different from those distributions obtained from pooling data across subjects (Fig. 5*B*) (two-sample Kolmogorov–Smirnov test, $p > 0.05$). The distributions from human data have a coefficient of variation (CV; ratio between SD and mean) close to 0.6, regardless of the fraction of dominance (Fig. 5*C*, blue dots) (slope not significantly different from zero, $p = 0.3$). Although the model shows a significant linear dependence on the fraction ($p < 0.05$), the dependence is weak, and the CV is consistently close to the experimental value (Fig. 5*C*, red dots). Importantly, the network predicts that the mean dominance durations of a percept should depend primarily on its fraction of dominance. The experimental data not only show this important qualitative feature but also follow quantitatively the idiosyncratic

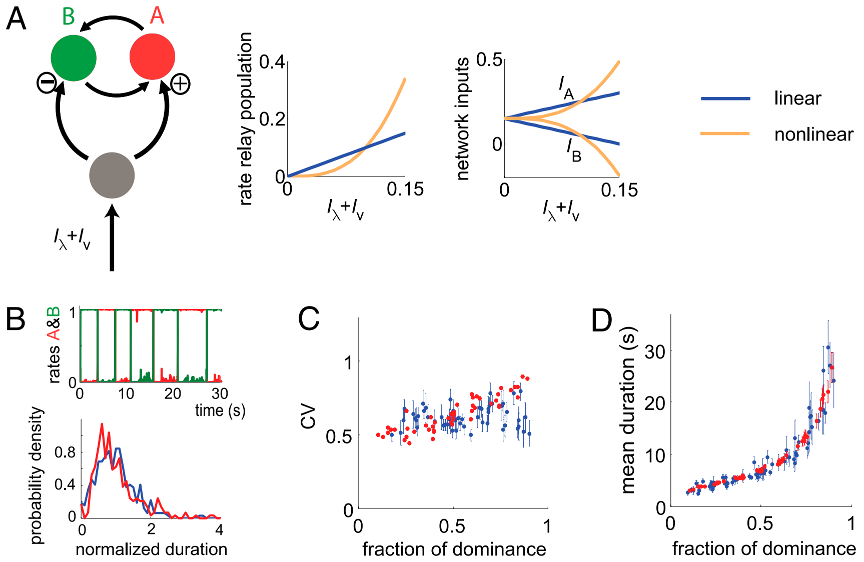

Fig. 5. Sampling and multiplicative rule in attractor neural networks. (*A*) Architecture of the network, linear, and nonlinear activation functions of the relay population and resulting inputs to the network. (*B*) Population firing rates as a function of time. (*Upper*) Red, population *A*; green, population *B*. (*Lower*) Distributions of dominance durations from the neural network model when the cues are neutral (red) and from the pooled data across subjects (blue) for the wavelength speed experiment in the neutral condition ($n = 320$). Time has been normalized so that the mean of the distributions is one. Because the distribution from the model corresponds to the case in which the cues are neutral (zero biasing currents), it is the same regardless of whether the activation function of the relay unit is linear or nonlinear. (*C*) CV of the dominance duration distribution of a percept as a function of its fraction of dominance for the data averaged across subjects (blue) and model (red). (*D*) Mean dominance duration of a percept as a function of its fraction of dominance for the experimental data averaged across subjects (blue) and for the model (red). Model error bars correspond to SEM across durations.

mean duration vs. fraction dependence obtained from the model (Fig. 5*D*). These results hold independently of whether the activation function of the relay population is linear (Fig. 5 *B*–*D*) or nonlinear (*SI Methods* and Fig. S2).

The slow dynamics of switches indicate that bistable perception generates temporally correlated samples (successive samples tend to be similar, which is indicated by the fact that percepts tend to linger for hundreds of milliseconds before switching), a property consistent with Langevin Monte Carlo sampling (28).

Therefore, the network generates a stochastic behavior consistent with bistable perception and makes nontrivial predictions about the dynamics of perceptual bistability. However, this behavior does not necessarily mean that the network follows the multiplicative rule. Interestingly, when the activation function in the relay population is nonlinear, the fractions of dominance do not combine multiplicatively (Figs. 2*D*, 3*D*, and 6*A*, orange dots). In contrast, when the activation function is linear-rectified, the network obeys the multiplicative rule (Figs. 2*D*, 3*D*, and 6*A*, blue dots). This result holds because the fraction of dominance time is a sigmoid function of the sum of input currents when the inputs to the network are linear (Fig. 6*B*, blue lines) but not when the inputs are nonlinear (Fig. 6*B*, orange lines). We show in *SI Methods* (Fig. S3) that these results hold even in a more realistic network with integrate and fire neurons.

Probabilistic Population Codes and Bayesian Sampling. We have shown in the previous sections how to build a recurrent network that implements the multiplicative rule, but we have not shown yet that the network samples the posterior distribution over image interpretations specified by the input signals. If the fraction of dominance for a given cue is the result of sampling the posterior distribution over image interpretations $p(s|I_i)$ (here $s = \{A,B\}$ and I_i is the current induced by cue $i = \{\lambda,v\}$), then the fraction of dominance and the posterior distribution should be the same function of the input current, I_i. Because the attractor network generates fractions of dominance that are sigmoid functions of the current (Eq. **3**), the attractor network is sampling the posterior distribution only if that distribution is also a sigmoid function of the input current, that is (Eq. **4**),

$$p(s=A\,|\,I_i)=f(s=A\,|\,I_i)=\frac{1}{1+e^{-2I_i/\sigma_{eff}^2}}. \qquad [\mathbf{4}]$$

Moreover, through Bayes rule, we know that (Eq. **5**)

$$p(s=A\,|\,I_i)\propto p(I_i\,|\,s=A), \qquad [\mathbf{5}]$$

where the function $p(I_i\,|\,s=A)$ corresponds to the variability in neural responses (in this case, one input current) over multiple presentations of the same stimulus s. Therefore, the key question is whether neural variability in vivo has a distribution consistent with Eqs. **4** and **5**. If this is not the case, attractor dynamics would not be sampling from the posterior distributions of s.

Experimentally, neural variability is typically assessed by measuring the variability in spike counts for a fixed s as opposed to the variability in input currents. Mapping input current onto spike counts is easy if we assume, as we did earlier, that the input current is proportional to the difference in spike counts vectors, $\mathbf{r}_A-\mathbf{r}_B$, from two presynaptic populations (e.g., V1 neurons with different depth and speed preferences) (37), one that prefers stimulus $s = A$ and the other that prefers stimulus $s = B$. One can then show (*SI Methods*) that Eqs. **4** and **5** are only satisfied when the distribution over either $\mathbf{r}_A$ or $\mathbf{r}_B$ given s takes the form $p(\mathbf{r}\,|\,s)\propto\phi(\mathbf{r})\exp(\mathbf{h}(s)\cdot\mathbf{r})$, where $\mathbf{h}(s)$ is a kernel related to the tuning curves and covariance matrix of the neural responses. Remarkably, this family of distributions, known as the exponential family with linear sufficient statistics, provides a very close approximation to the variability observed in vivo (22, 38).

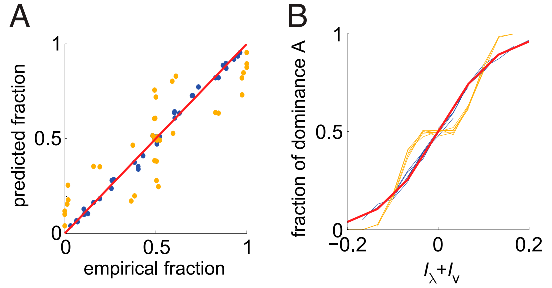

Fig. 6. (*A*) Predicted fractions from the multiplicative rule vs. observed fractions of dominance generated by the neural network with nonlinear (orange) and linear (blue) inputs (*SI Methods*). As observed with the energy model (Fig. 4*E*), the network follows the multiplicative rule only when the relay population has a linear activation function. (*B*) Fraction of dominance of state *A* as a function of the total input when the relay population is nonlinear (orange) and linear (blue). The latter are well-fitted by a sigmoid function (red), which was the case with the energy model (Fig. 4*F*).

This family of distributions corresponds also to a form of neural code known as probabilistic population codes (22). In other words, our results show that attractor dynamics can be used to sample from a posterior distribution encoded by a probabilistic population code using the exponential family with linear sufficient statistics.

Discussion

We have reported that the fraction of dominance in bistable perception behaves as a probability. This result supports the notion that the visual system samples the posterior distribution over image interpretations. In addition, we showed that attractor networks can implement Bayesian sampling only when the variability of neuronal activity follows the exponential family with linear sufficient statistics, as observed experimentally.

This last result is important, but using the exponential family has another advantage. Several works have reported that humans perform near-optimal cue integration in a variety of settings (1–8). It is, therefore, essential that the combination of inputs that leads to the multiplicative rule in an attractor network also results in optimal cue integration. We saw that inputs need to be added to observe the multiplicative rule in an attractor network. Adding two inputs does not necessarily result in optimal cue integration, but again, when the variability of cortical activity follows the exponential family with linear sufficient statistics, it is the optimal combination rule for cue integration (22). Therefore, the fact that the neural variability follows the exponential family allows both Bayesian sampling and optimal integration of evidence with attractor networks.

Our study is not the first study to investigate cue combination and perceptual bistability, but previous works did not test whether bistable perception is akin to what we defined as Bayesian sampling (19, 20). The fact that bistable perception alternates between two interpretations is certainly suggestive of a sampling process but not necessarily of Bayesian sampling. For instance, the orange dots in Fig. 6*A* show an example of a network that stochastically oscillates with gamma-like distributions over percept durations (Fig. 5*B*), as observed in our experimental data. The kind of analysis that has been used in previous studies to argue that bistable perception is a form of sampling (19, 20) would also conclude that this network is sampling. However, this particular network does not perform Bayesian sampling; it does not follow the multiplicative rule (Fig. 6*A*). In contrast, our experimental results make it clear that bistable perception follows the multiplicative rule predicted by Bayesian sampling.

Bayesian sampling has several computational advantages. For instance, in the context of reinforcement learning, when the statistics of the world is fixed, the optimal solution involves picking the action that is the most likely to be rewarded; however, when the statistics of the world change over the time, sampling from the posterior distribution, which is a form of exploratory behavior (21, 39), is more sensible (40). Interestingly, bistable perception implements a form of sampling that could be used to smoothly interpolate between pure exploration (sampling from the posterior) and pure exploitation (choosing the action that is the most likely to be rewarded). Indeed, our results suggest that bistable perception samples from posterior distributions that are raised to a power, p^n, where n can take any value (*SI Methods*). When n is large, the most likely state is sampled on almost every iteration, which corresponds to exploitation, whereas setting n close to zero leads to exploratory behavior.

The fact that low-level vision and perhaps low-level perception might involve sampling is particularly interesting in light of several other recent findings suggesting that higher-level cognitive tasks, like causal reasoning (41, 42) and decision-making (43), might also involve some form of sampling. Sampling may turn out to be a general algorithm for probabilistic inference in all domains.

Methods

Experimental Methods. The stimulus consisted of two superimposed square-wave gratings, denoted α and β, moving at an angle of 160° between their directions of motion behind a circular aperture (21) (Fig. 1*A*) with the parameters specified in *SI Methods*. The gratings consisted of gray bars of equal luminance presented on a white background. Where the gray bars intersected, the luminance was set to that of the bars (as if one of the bars was occluding the other bar). Observers were asked to continually report their percept by holding down one of two designated keys [i.e., motion direction (right or left) of the grating that they perceived as being behind the other grating] and not to press any key if they were not certain. We measured, in each trial, the accumulated time that either percept (i.e., depth ordering) was dominant and computed the fraction of time that percept $s = \{A,B\}$ dominated as $f(s)$ = (the cumulative time percept s was reported as dominant)/(the total time that either of the percepts was reported as dominant). Therefore, this fraction corresponds to the proportion of time that percept s dominated. Percept A denotes the percept in which grating α is behind grating β (and conversely, percept B). Fractions of dominance shown in the figures correspond to averaged values of the fractions across trials and observers, and error bars correspond to SEM across the population.

Mathematical Methods. The derivations of the multiplicative rule and stronger cue take all rule and the descriptions of the energy, rate-based, and spiking models are presented in *SI Methods*.

ACKNOWLEDGMENTS. We thank Jan Drugowitsch and Robbie Jacobs for their suggestions and comments. We are also very grateful to Thomas Thomas and Bo Hu for their assistance during the experimental setup and Vick Rao for his help in using the cluster. D.C.K. is supported by National Institutes of Health Grant EY017939. A.P. is supported by National Science Foundation Grant BCS0446730 and the Multidisciplinary University Research Initiative (MURI) Grant N00014-07-1-0937. This work was also partially supported by National Eye Institute Award P30 EY001319.

1. Ernst MO, Banks MS (2002) Humans integrate visual and haptic information in a statistically optimal fashion. *Nature* 415:429–433.
2. Jacobs RA (1999) Optimal integration of texture and motion cues to depth. *Vision Res* 39:3621–3629.
3. Landy MS, Maloney LT, Johnston EB, Young M (1995) Measurement and modeling of depth cue combination: In defense of weak fusion. *Vision Res* 35:389–412.
4. van Beers RJ, Sittig AC, Gon JJ (1999) Integration of proprioceptive and visual position-information: An experimentally supported model. *J Neurophysiol* 81:1355–1364.
5. Körding KP, Wolpert DM (2006) Bayesian decision theory in sensorimotor control. *Trends Cogn Sci* 10:319–326.
6. Hillis JM, Watt SJ, Landy MS, Banks MS (2004) Slant from texture and disparity cues: Optimal cue combination. *J Vis* 4:967–992.
7. Knill DC (2007) Robust cue integration: A Bayesian model and evidence from cue-conflict studies with stereoscopic and figure cues to slant. *J Vis* 7:5.1–5.24.
8. Knill DC (2003) Mixture models and the probabilistic structure of depth cues. *Vision Res* 43:831–854.
9. Tjan BS, Braje WL, Legge GE, Kersten D (1995) Human efficiency for recognizing 3-D objects in luminance noise. *Vision Res* 35:3053–3069.
10. Gold JM, Tadin D, Cook SC, Blake R (2008) The efficiency of biological motion perception. *Percept Psychophys* 70:88–95.
11. Kersten D, Mamassian P, Yuille A (2004) Object perception as Bayesian inference. *Annu Rev Psychol* 55:271–304.
12. Hinton GE (2007) Learning multiple layers of representation. *Trends Cogn Sci* 11: 428–434.
13. Fiser J, Berkes P, Orbán G, Lengyel M (2010) Statistically optimal perception and learning: From behavior to neural representations. *Trends Cogn Sci* 14:119–130.
14. Blake R (2001) A primer on binocular rivalry. *Brain and Mind* 2:5–38.
15. Blake R, Logothetis NK (2002) Visual competition. *Nat Rev Neurosci* 3:13–21.
16. Dayan P (1998) A hierarchical model of binocular rivalry. *Neural Comput* 10:1119–1135.
17. Necker LA (1832) Observations on some remarkable phenomenon which occurs on viewing a figure of a crystal of geometrical solid. *Lond Edinburgh Phil Mag J Sci* 3: 329–337.
18. Rubin E (1958) Figure and ground. *Readings in Perception*, eds Beardslee DC, Wertheimer M (Van Nostrand Reinhold, New York), pp 194–203.
19. Sundareswara R, Schrater PR (2008) Perceptual multistability predicted by search model for Bayesian decisions. *J Vis* 8:12.1–12.19.
20. Hoyer PO, Hyvarinen A (2003) Interpreting neural response variability as Monte Carlo sampling of the posterior. In Becker S, et al., editors. Advances in Neural Information Processing Systems 15. MIT Press; 2003. pp. 277–284.
21. Moreno-Bote R, Shpiro A, Rinzel J, Rubin N (2008) Bi-stable depth ordering of superimposed moving gratings. *J Vis* 8:20.1–20.13.
22. Ma WJ, Beck JM, Latham PE, Pouget A (2006) Bayesian inference with probabilistic population codes. *Nat Neurosci* 9:1432–1438.
23. Mamassian P, Landy MS (1998) Observer biases in the 3D interpretation of line drawings. *Vision Res* 38:2817–2832.
24. Mamassian P, Landy MS (2001) Interaction of visual prior constraints. *Vision Res* 41: 2653–2668.
25. Priebe NJ, Mechler F, Carandini M, Ferster D (2004) The contribution of spike threshold to the dichotomy of cortical simple and complex cells. *Nat Neurosci* 7:1113–1122.
26. Carandini M, Ferster D (2000) Membrane potential and firing rate in cat primary visual cortex. *J Neurosci* 20:470–484.
27. Moreno-Bote R, Rinzel J, Rubin N (2007) Noise-induced alternations in an attractor network model of perceptual bistability. *J Neurophysiol* 98:1125–1139.
28. Bishop CM (2006) *Pattern Recognition and Machine Learning* (Springer, Berlin).
29. Sheinberg DL, Logothetis NK (1997) The role of temporal cortical areas in perceptual organization. *Proc Natl Acad Sci USA* 94:3408–3413.
30. Shpiro A, Moreno-Bote R, Rubin N, Rinzel J (2009) Balance between noise and adaptation in competition models of perceptual bistability. *J Comput Neurosci* 27:37–54.
31. Laing CR, Chow CC (2002) A spiking neuron model for binocular rivalry. *J Comput Neurosci* 12:39–53.
32. Markram H, Tsodyks M (1996) Redistribution of synaptic efficacy between neocortical pyramidal neurons. *Nature* 382:807–810.
33. Abbott LF, Varela JA, Sen K, Nelson SB (1997) Synaptic depression and cortical gain control. *Science* 275:220–224.
34. Leopold DA, Logothetis NK (1996) Activity changes in early visual cortex reflect monkeys' percepts during binocular rivalry. *Nature* 379:549–553.
35. Levelt WJM (1968) *On Binocular Rivalry* (Mouton, The Hague).
36. Hupé JM, Rubin N (2003) The dynamics of bi-stable alternation in ambiguous motion displays: A fresh look at plaids. *Vision Res* 43:531–548.
37. Cumming BG, DeAngelis GC (2001) The physiology of stereopsis. *Annu Rev Neurosci* 24:203–238.
38. Graf AB, Kohn A, Jazayeri M, Movshon JA (2011) Decoding the activity of neuronal populations in macaque primary visual cortex. *Nat Neurosci* 14:239–245.
39. Moreno-Bote R, Shpiro A, Rinzel J, Rubin N (2010) Alternation rate in perceptual bistability is maximal at and symmetric around equi-dominance. *J Vis* 10(11):1,1–18.
40. Sutton RS, Barto AG (1998) Reinforcement learning: An introduction. *Adaptive Computation and Machine Learning* (MIT Press, Cambridge, MA).
41. Griffiths TL, Tenenbaum JB (2005) Structure and strength in causal induction. *Cognit Psychol* 51:334–384.
42. Tenenbaum JB, Griffiths TL, Kemp C (2006) Theory-based Bayesian models of inductive learning and reasoning. *Trends Cogn Sci* 10:309–318.
43. Sugrue LP, Corrado GS, Newsome WT (2004) Matching behavior and the representation of value in the parietal cortex. *Science* 304:1782–1787.

Supporting Information

Moreno-Bote et al. 10.1073/pnas.1101430108

SI Methods

Experimental Methods. ***Observers.*** A total of seven naïve observers participated in the experiments: four in experiment 1 (observers 1–4; two females) and four in experiment 2 (observers 4–7; two females). All observers had normal or corrected to normal vision. They were paid $10 per session for their participation, and they provided informed consent according to the guidelines of the University of Rochester Research Subjects Review Board.

Stimulus. The stimulus consisted of two superimposed square-wave gratings, denoted α and β, moving at an angle of 160° between their directions of motion (±80 from the vertical degree) behind a circular aperture (1), which is shown schematically in Fig. 1*A*. The luminance of the intersections between the gratings was identical to that in the bars, and the values are specified below. The duty cycle of the gratings was fixed at 0.2 (duty cycle = bar width/wavelength). The diameter of the aperture was 14°. Luminance outside the aperture was 9 cd/m^2. A circular fixation point (radius = 0.2°, luminance = 26 cd/m^2) was overlaid on a small homogeneous circular region (radius = 1°, luminance = 0.3 cd/m^2) that covered the center of the display. Observers sat at a distance 57 cm from the screen.

Experiment 1 (wavelength and speed cues). For the congruent cues (C) condition, grating β had a fixed wavelength, $\lambda = 3°$, and speed, $v = 3°/s$ (neutral values), whereas the wavelength and speed of grating α took one of the following values in each trial: $\lambda = 1.9°$, 2.3°, and 3° and $v = 3°$, 6°, and 10°/s. For the incongruent cues (IC) condition, grating β had a fixed wavelength, $\lambda = 3°$ (neutral value), and its speed took the values $v = 3°$, 6°, and 10°/s; grating α had a fixed speed, $v = 3°/s$ (neutral value), whereas its wavelength took the values λ=1.9°, 2.3°, and 3°. The bars of the gratings had luminance 21 cd/m^2 and were presented on a white background (38 cd/m^2).

Experiment 2 (wavelength and disparity cues). For the congruent cues (C) condition, grating β had a fixed wavelength, $\lambda =$ 3°, and speed, $v = 6°/s$ (neutral values), whereas the wavelength and (uncrossed) disparity of grating α took one of the following values in each trial: $\lambda = 1.9°$, 2.3°, and 3° and $d = 0$, 2.4, and 4.8 arcmin. For the incongruent cues (IC) condition, grating β had a fixed wavelength, $\lambda = 3°$ (neutral value), whereas its (uncrossed) disparity took the values $d = 0$, 2.4, and 4.8 arcmin; grating α had a fixed speed, $v = 3°$ (neutral value), whereas its wavelength took the values $\lambda = 1.9°$, 2.3°, and 3°. Gratings were displayed in red (5 cd/m^2) over a black background (0.1 cd/m^2).

Apparatus. The stimuli were generated by an Intel-based PC running a C program and using the OpenGL graphics library, and they were displayed on a 20-in cathode ray tube (CRT) screen. The stimuli in experiment 1 were displayed at 75 Hz with a resolution of 1,280 × 1,024 pixels. Shutter stereo glasses were used in experiment 2; the stimuli were displayed at 120 Hz (60 Hz per eye) with a resolution of 1,152 × 864 pixels.

Experimental procedure. Observers sat in front of a computer screen with their heads supported by a chinrest. They were asked to continually report their percept by holding down one of two designated keys [i.e., motion direction (right or left) of the grating that they perceived as being behind the other grating]. Observers were given passive viewing instructions (not to try to perceive one possibility more than the other possibility and just to report the spontaneous changes), and they were instructed not to press either key if the percept was unclear. Observers fixated the central spot during the whole 1-min duration of each trial. All combinations of wavelength and speed or wavelength and disparity were used in a randomized order, and each combination was repeated four times. The global directions of motion of the two gratings were randomized (up-right, up-left, down-right, and down-left; always ±80° from the vertical and the global direction of motion did not produce any significant effect). Observers ran a total of 36 trials of 1 min each in a single session; they were instructed to take a 10- to 30-s rest between the trials. Subjects repeated the same experiment in two or three sessions to equalize size of the error bars across subjects.

Long presentations (∼1 min) were chosen over short presentations (∼1 s) (2), because the latter is known to have strong saturating effects (confirmed by pilot experiments) (3), which would have made it more difficult to find a large range of parameters for which the fractions of dominances were different from both zero and one for all subjects.

Analysis. On each trial, we measured the accumulated time that each percept (i.e., depth ordering) was dominant and computed the fraction of time that percept s ($s = \{A,B\}$) dominated. This fraction was denoted $f(s)$ and defined as the ratio of the cumulative time that percept s was reported as dominant over the total time that either of the percepts was reported as dominant. Percept A denotes the percept in which grating α is behind grating β (and conversely, percept B). This fraction is a number between zero and one, with a value of 0.5 indicating that the two possible percepts were equally likely. The fractions are first averaged across orientations, because this variable was found to have no impact on the fraction of dominance. Fractions of dominance shown in the figures correspond to averaged values of the fractions across observers, and error bars correspond to SEMs across the population if not stated otherwise.

Predictions for the Fractions of Dominance Time. ***Multiplicative rule.*** In the text, we have described the multiplicative rule (Eq. **1**) and the stronger cue takes all rule. A detailed derivation of the multiplicative rule is given here. In this section, $f_{\lambda s}(s) = f(s \mid \Delta\lambda, \Delta v)$ denotes the measured fraction of dominance of each stimulus interpretation s ($s = A$ or B corresponding to the two possible depth orderings) when the cues take values $\Delta\lambda$ and Δv (Δv should be replaced by the differences in disparity in the second experiment). According to the sampling hypothesis, this fraction is proportional to the probability (or some power of it) of the percept given the sensory evidence, that is, $f_{\lambda v}(s) \propto p^n(s \mid \Delta\lambda, \Delta v)$. Note that this relationship does not assume any specific dynamics for the sampling process of the posterior and also allows for tempered sampling of the probability distribution when the power n is different from one (4). Assuming that the values of the cues are conditionally independent given s and using Bayes' rule two times, we obtain (**S1**)

$$\begin{aligned} f_{\lambda v}(s) &\propto p^n(s \mid \Delta\lambda, \Delta v) \\ &\propto p^n(\Delta\lambda, \Delta v \mid s) p^n(s) \\ &\propto p^n(\Delta\lambda \mid s) p^n(\Delta v \mid s) p^n(s) \\ &\propto \frac{p^n(s \mid \Delta\lambda) p^n(s \mid \Delta v)}{p^n(s)}, \end{aligned} \tag{S1}$$

where $p(s)$ is the prior over depth orderings. Likewise, we can define the fractions of dominance of percept s when we manipulate individual cues while keeping the other cue constant. Let us fix first the speed at $\Delta v = \Delta v_0$ (where Δv_0 is an arbitrary reference point) and manipulate the relative wavelengths of the gratings $\Delta\lambda$. Using expression **S1** for this particular speed difference leads to (**S2**)

$$f_\lambda(s) \propto p^n(s \mid \Delta\lambda, \Delta v_0) \propto \frac{p^n(s \mid \Delta\lambda) p^n(s \mid \Delta v_0)}{p^n(s)}. \quad \textbf{[S2]}$$

Similarly, if we fix the relative wavelengths at $\Delta\lambda = \Delta\lambda_0$, we find that the fraction of dominance for the speed difference Δv is given by (Eq. S3)

$$f_v(s) \propto p^n(s \mid \Delta\lambda_0, \Delta v) \propto \frac{p^n(s \mid \Delta\lambda_0) p^n(s \mid \Delta v)}{p^n(s)}. \quad \textbf{[S3]}$$

We can now insert $p^n(s \mid \Delta\lambda)$ from Eq. S2 and $p^n(s \mid \Delta v)$ from Eq. S3 into Eq. S1 to obtain (Eq. S4)

$$f_{\lambda v}(s) \propto f_\lambda(s) \times f_v(s) \times \frac{p^n(s)}{p^n(s \mid \Delta\lambda_0) p^n(s \mid \Delta v_0)}. \quad \textbf{[S4]}$$

Finally, we note that if we set the reference differences in speed and wavelength, Δv_0 and $\Delta\lambda_0$, to zero (i.e., their neutral values), then $p(s), p(s \mid \Delta v_0 = 0), p(s \mid \Delta\lambda_0 = 0)$ should all be equal to one-half for symmetry reasons, which leads to (Eq. S5)

$$f_{\lambda v}(s) \propto f_\lambda(s) \times f_v(s). \quad \textbf{[S5]}$$

After normalized, this equation is equivalent to the multiplicative rule (Eq. **1**). This relation holds whether subjects are sampling from posterior distributions on depth or from posterior distributions raised to an arbitrary power n. Therefore, our experimental results are consistent with a sampling process of an underlying posterior distribution over depth orderings raised to an arbitrary power. Our technique does not specify the power to which the probability has been raised, because a probability raised to a power behaves exactly like any other probability distribution (i.e., it gets combined according to the rules of probability).

Combination rule over continuous variables. In the previous section, we assumed that subjects estimate depth ordering, which is to say that they infer the value of a binary variable from the sensory information. It is possible, however, that subjects first compute a posterior distribution over the depth of the gratings, which is a continuous variable, and then recover a probability distribution over depth ordering through integration. This approach does not yield the same multiplicative rule as before. Instead, as we show below, we obtain (Eq. S6)

$$f_{\lambda v} = \Phi(\Phi^{-1}(f_\lambda) + \Phi^{-1}(f_v)), \quad \textbf{[S6]}$$

where Φ is the cumulative of a normal distribution and Φ^{-1} is its inverse. Although this rule looks quite different from the multiplicative rule, it leads to nearly identical predictions (Fig. S4).

To derive this rule, we assume that (*i*) on any given trial, the subject has available noisy sensory estimates of $\Delta\lambda$, Δv, $\Delta\widehat{\lambda}$, and $\Delta\widehat{v}$, (*ii*) the likelihood functions on the difference in depth associated with each of the sensory measurements is Gaussian with means $\hat{z}_\lambda$ and $\hat{z}_v$ and variances σ_λ^2 and σ_v^2, and (*iii*) the prior on the difference in depth, z, is uniform. The posterior density function on z is then Gaussian with mean $w_\lambda \hat{z}_\lambda + w_v \hat{z}_v$ and variance $w_\lambda^2 \sigma_\lambda^2 + w_v^2 \sigma_v^2$, where $w_\lambda = R_\lambda/(R_\lambda + R_v)$, $w_v = R_v/(R_\lambda + R_v)$, and the R_i terms represent the reliabilities of the two cues given by $R_i = 1/\sigma_i^2$. For an observer that samples from this posterior, the frequency of seeing a difference in depth greater than zero is given by (Eq. S7)

$$f_{\hat{\lambda}\hat{v}} = p(z > 0 \mid \Delta\widehat{\lambda}, \Delta\widehat{v}) = \Phi\left(\frac{1}{\sqrt{R_\lambda + R_v}}(R_\lambda \hat{z}_\lambda + R_v \hat{z}_v)\right). \quad \textbf{[S7]}$$

We do not have this frequency available psychophysically, because it is conditioned on particular noisy sensory estimates of $\Delta\lambda$ and Δv. Rather, we have frequencies averaged across trials; thus, we have available the average frequency with which subjects report one percept over another from trial to trial. To derive an expression, we replace the sensory signals $\Delta\widehat{\lambda}$ and $\Delta\widehat{v}$ with a representation of the depth difference indicated by the sensory signals, $\widehat{z} = w_\lambda \hat{z}_\lambda + w_v \hat{z}_v$, where $\hat{z}_\lambda$ and $\hat{z}_v$ are the means of the likelihoods associated with $\Delta\widehat{\lambda}$ and $\Delta\widehat{v}$, respectively, and the weights are as described above. Noticing that we can write $p(z > 0 \mid \Delta\widehat{\lambda}, \Delta\widehat{v})$ as $p(z > 0 \mid \widehat{z})$, the average frequency of a positive depth difference interpretation, $f_{\lambda v}$, is given by (Eq. S8)

$$f_{\lambda v} = \int_{-\infty}^{\infty} p(z > 0 \mid \widehat{z}) p(\widehat{z} \mid z_\lambda, z_v) d\widehat{z} = \int_0^{\infty} \int_{-\infty}^{\infty} p(z \mid \widehat{z}) p(\widehat{z} \mid z_\lambda, z_v) dz d\widehat{z}, \quad \textbf{[S8]}$$

where $p(\widehat{z} \mid z_\lambda, z_v)$ represents the trial to trial distribution of depth differences derived from the values of the two cues. This distribution is Gaussian with mean $w_\lambda z_\lambda + w_v z_v$ and variance $w_\lambda^2 \sigma_\lambda^2 + w_v^2 \sigma_v^2$ with the weights as defined above. We can, therefore, rewrite Eq. S8 as (Eq. S9)

$$f_{\lambda v} = \int_0^{\infty} \int_{-\infty}^{\infty} f\left(z - \widehat{z}; 0, \frac{1}{R_\lambda + R_v}\right) f\left(\widehat{z}; w_\lambda z_\lambda + w_v z_v, \frac{1}{R_\lambda + R_v}\right) dz d\widehat{z}, \quad \textbf{[S9]}$$

where $f(z; \mu, \sigma^2)$ is a Gaussian with mean μ and variance σ^2. The inner integral is simply a convolution of two Gaussians with the same variance, resulting in (Eq. S10)

$$f_{\lambda v} = \int_0^{\infty} f\left(z; w_\lambda z_\lambda + w_v z_v, \frac{2}{R_\lambda + R_v}\right) dz = \Phi\left(\sqrt{\frac{1}{2(R_\lambda + R_v)}}(R_\lambda z_\lambda + R_v z_v)\right). \quad \textbf{[S10]}$$

Note that the cues are neutral when $\Delta\lambda = 0$ and $\Delta v = 0$, because then, $f_{\lambda v} = \Phi(0) = 1/2$. This scenario corresponds to a situation in which both $z_\lambda = 0$ and $z_v = 0$. When the speed cue is neutral, $z_v = 0$, and the percept corresponding to $z > 0$ will be reported with a frequency $f_\lambda = \Phi\left(\sqrt{\frac{1}{2(R_\lambda + R_v)}}(R_\lambda z_\lambda)\right)$. When the wavelength cue is neutral, $z_\lambda = 0$, and the percept corresponding to $z > 0$ will be reported with a frequency $f_v = \Phi\left(\sqrt{\frac{1}{2(R_\lambda + R_v)}}(R_v z_v)\right)$. Thus, we can rewrite Eq. S10 as (Eq. S11)

$$f_{\lambda v} = \Phi(\Phi^{-1}(f_\lambda) + \Phi^{-1}(f_v)). \quad \textbf{[S11]}$$

The above derivation assumes that the sensory signals are corrupted by a constant noise term on each trial. The same results hold if we assume time-varying noise within a trial. Furthermore, Eq. S11 is valid even when the observer has inaccurate estimates of the cue variance, because it only changes the weights and variances of the Gaussian inside the integral in Eq. S10. In this case, Eq. S9 becomes (Eq. S12)

$$f_{\lambda v} = \int_{0}^{\infty} \int_{-\infty}^{\infty} f(z-\widehat{z}; 0, \sigma_z^2) f(\widehat{z}; w_\lambda z_\lambda + w_v z_v, w_\lambda \sigma_\lambda^2 + w_\lambda \sigma_v^2) dz d\widehat{z}$$
$$= \Phi\left(\frac{1}{\sqrt{\sigma_z^2 + w_\lambda^2 \sigma_\lambda^2 + w_v^2 \sigma_v^2}} (w_\lambda z_\lambda + w_v z_v) \right), \qquad \textbf{[S12]}$$

where σ_z^2 is the variance associated with the posterior density function on z computed using the incorrect estimates of signal variance. The second term in the integral contains the true signal variances, because it represents the variability in the depth difference signaled by different noisy measurements of wavelength and speed. This result is analogous to the invariance of the multiplicative rule to the power-law transformations of the posterior probabilities in the discrete case—incorrect estimates of cue variance are equivalent to raising the Gaussian likelihood functions to arbitrary powers. The implication of this result is that the multiplicative rule and Eq. **S11** do not require that the brain combines cues optimally but simply that it uses a weighted average.

Attractor Neural Networks. ***Energy model.*** We model the dynamics of two populations, A and B, whose states are described by their firing rates r_A and r_B, respectively. In the energy model, the variable difference between firing rates, $r = r_A - r_B$, obeys (Eq. **S13**)

$$\tau \frac{d}{dt} r = -4r(r^2 - 1) + g(I_\lambda, I_v) + n(t) \qquad \textbf{[S13]}$$

(Eq. **2**), where $n(t)$ is as an Ornstein–Uhlenbeck process (5) with zero mean and deviation σ ($\sigma = 1.2\text{s}^{-1/2}$) (Eq. **S14**):

$$\frac{d}{dt} n_i = -\frac{n_i}{\tau_s} + \sigma \sqrt{\frac{2}{\tau_s}} \xi_i(t). \qquad \textbf{[S14]}$$

Here, $\tau_s = 100$ ms, and $\xi_i(t)$ is a white noise process with zero mean and unit variance, $\langle \xi_i(t) \xi_i(t') \rangle = \delta(t-t')$, where $\delta(t-t')$ is the δ-function. The input-dependent drift takes the form $g(I_\lambda, I_v) = I_\lambda + I_v + \varepsilon(I_\lambda I_v^2 + I_v^2 I_\lambda)$, where ε weights the strength of the nonlinearity and I_λ and I_v are the currents supporting dominance of population A vs. population B from the wavelength and speed cues, respectively. This function corresponds to a simple expansion in powers of the currents of an odd function up to third order. Including the pure third-order terms I_λ^3 and I_v^3 does not qualitatively change the results, because their effects on the fractions of dominance can be reabsorbed in the linear terms.

Eq. **3** can be derived analytically from the energy model by computing the mean dominance duration of each percept. This quantity corresponds to the mean escape time from one of the potential wells. As is well-known from the theory of first-passage times, the mean escape time from a well depends exponentially on its energy barrier (i.e., vertical distance from the minimum to the local maximum) for small noise variances (5, 6), the case of interest in our problem. Therefore, the fraction of dominance of percept A, which is proportional to its mean dominance duration, is itself proportional to the exponential of the energy barrier (**S15**),

$$f \propto T(A) \propto e^{(E_0 + \Delta E(A))/\sigma_{eff}^2} \propto e^{\Delta E(A)/\sigma_{eff}^2}, \qquad \textbf{[S15]}$$

where σ_{eff}^2 is the effective noise in the system, E_0 is the energy barrier when the input currents are equal to zero (i.e., $I_\lambda = I_s = 0$), and $\Delta E(A)$ is the change in the energy barrier induced by nonzero currents. Eq. **3** can be obtained from expression **S15** by noting that, for the linear model, the change of the energy barrier is linear in the sum of the currents, $\Delta E(A) = c(I_\lambda + I_v)$, with c being a constant close to one for small current values.

Rate-based neural model. In the rate-based network, the firing rate r_i for each population i ($i = A,B$) follows the coupled differential equations (Eqs. **S16** and **S17**)

$$\begin{aligned} \tau \frac{d}{dt} r_A &= -r_A + R(w_{exc} r_A - w_{inh} d_B r_B + I_0 + r_{relay}(I_\lambda, I_v) + n_A) \\ \tau \frac{d}{dt} r_B &= -r_B + R(w_{exc} r_B - w_{inh} d_A r_A + I_0 - r_{relay}(I_\lambda, I_v) + n_B) \end{aligned}, \qquad \textbf{[S16 and S17]}$$

where $I_0 = 0.15$ is a constant background current. The rate of the relay population is (Eq. **S18**)

$$r_{relay}(I_\lambda, I_v) = (1-\varepsilon)(I_\lambda + I_v) + \varepsilon C (I_\lambda + I_v)^3, \qquad \textbf{[S18]}$$

where the parameter ε specifies the degree of nonlinearity in the inputs currents (in Figs. 5 and 6, ε was set to zero for the linear relay population and set to one for the nonlinear relay population) and $C = 100$ is a normalization constant ensuring that the two terms in the sum have roughly the same magnitude for the values of the currents used in the stimulations (range between 0 and 0.1). The choice of the cubic nonlinearity is consistent with the experimentally observed input to rate transfer functions in primary visual cortex (7).

The above equations and the architecture described in Fig. 5*A* correspond to the case in which the sum of the currents is positive; when the sum is negative, the firing rate of the relay population is zero (i.e., the sum of the input currents are rectified). We assume that there is an additional relay population that is active when the sum of the currents is negative and delivers inputs to the network with a firing rate equal to $-r_{relay}(I_\lambda, I_v)$, which is positive. The connectivity of this additional relay population was assumed to be the reverse of the one shown in Fig. 5*A*. The firing rates of the populations relax to their steady-state value with time constant $\tau = 10$ ms. The function $R(x)$ is the input current to firing rate transfer function taken to be a sigmoid $R(x) = 1/(1 + e^{-x/k})$, with $k = 0.2$. Recurrent excitatory connections in each population with strength $w_{\text{exc}} = 1$ provide positive feedback. Cross-inhibition between the population has strength $w_{\text{inh}} = 2$ and generates winner take all behavior (i.e., only one population has large activity at any time). Synaptic depression is modeled as a multiplicative term, d_i, that lowers the effective inhibition exerted from one population to the competing population and follows the equation $\tau_d \frac{d}{dt} d_i = 1 - d_i - u r_i d_i$, where $\tau_d = 2$ s is the time scale of depression and u determines how quickly vesicles are depleted by presynaptic activity ($u = 0.6$) (8, 9). The populations receive independent fluctuating currents, $n_A(t)$ and $n_B(t)$, modeled as in Eq. **S14** with intensity $\sigma = 0.24\text{s}^{-1/2}$.

Fits from the rate-based neuronal model to experimental data. The rate-based model generates predictions about the fractions of dominance for each percept that one should observe when the two cues are nonneutral given the fractions observed when only one cue was nonneutral. The blue dots in Figs. 2*D*, 3*D*, and 6*A* show that, when the relay population has a linear activation function, the network generates fractions of dominance for the two nonneutral cues condition that matches the multiplicative rule (Eq. **1**), whereas the orange dots in Figs. 2*D*, 3*D*, and 6*A* show that, when the relay population was nonlinear, the generated fractions did not follow the multiplicative rule.

To generate the predictions from the rate-based model, we first fitted the single nonneutral cue conditions as follows. In the linear version of the system, the fractions of dominance of percept A follow a sigmoid function of the current (Fig. 6*B*, blue lines), that was fitted by a sigmoid (logistic) function (Fig. 6*B*, red line) (Eq. **S19**)

$$f_i(s=A)=\frac{1}{1+e^{-2I_i/\sigma^2_{best}}}, \qquad \textbf{[S19]}$$

where σ^2_{best} is the best parameter estimate and I_i is the current. For each data point in the experimental set, we found the corresponding current I_i that gave the observed fraction f_i in the single nonneutral cue condition ($i=\{\lambda,v\}$). Finally, the prediction of the fraction of dominance that should be observed when the two cues were nonneutral was generated by using the previously fitted sigmoid functions now applied to the sum of the currents, that is (Eq. **S20**),

$$f_{\lambda v}(s=A)=\frac{1}{1+e^{-2(I_\lambda+I_v)/\sigma^2_{best}}}. \qquad \textbf{[S20]}$$

It is easy to show that this prediction is equivalent to applying the multiplicative rule directly to the fractions of dominance in the single nonneutral cue condition.

For the nonlinear version of the network, the fit was as follows; instead of using a sigmoid function, we fitted the fractions of dominance in the single nonneutral cue condition in Fig. 6*B* (orange lines) to the function (Eq. **S21**)

$$f_i(s=A)=\frac{1}{1+e^{-2I_i^3/\sigma^2_{best}}}. \qquad \textbf{[S21]}$$

From this fit, we can compute again the corresponding current I_i that gives the observed fraction f_i in the single nonneutral cue condition. Finally, we generated predictions for the fraction in the two nonneutral cues condition from this model by using (Eq. **S22**)

$$f_{\lambda v}(s=A)=\frac{1}{1+e^{-2(I_\lambda^3+I_v^3)/\sigma^2_{best}}}. \qquad \textbf{[S22]}$$

This equation led to predicted fractions of dominance that substantially deviated from the multiplicative rule and hence, from the data (Figs. 2*D* and 3*D*, orange dots).

Spiking attractor network. Results. We also built a network of integrate and fire neurons that generates fractions of dominance that follow the rules of probabilistic inference (Fig. S3). The architecture of this network is similar to that described in Fig. 5*A* but with the additional ingredient that there is a population of inhibitory neurons associated with each excitatory population (Fig. S3*A*). Fig. S3*B* shows a time series of dominances of the two excitatory populations (Fig. S3*B Upper* displays population firing rates) and the raster plots (Fig. S3*B Lower*) indicating the spike timings of every neuron in the network (population A runs from neurons 1 to 50, with the last 10 neurons corresponding to the A inhibitory population, whereas population B runs from neurons 51 to 100, with the last 10 neurons corresponding to the inhibitory subpopulation). The network generates standard stochastic behavior with a distribution of dominance durations that has a coefficient of variation of 0.54 (Fig. S3*C*). When the spiking network receives inputs from a linear relay population, the network generates fractions of dominance that follow the multiplicative rule (Fig. S3*D*, blue dots), whereas when the relay population is nonlinear, the fractions do not combine multiplicatively (Fig. S3*D*, orange dots). As required for sampling, the fractions are well-described by a sigmoid function when the relay population is linear but not when it is nonlinear (Fig. S3*E*).

Network description. Each neuronal population contains $n=50$ leaky integrate and fire neuron models (20% of them are inhibitory and the remaining are excitatory). Coupling is with instantaneous current injections and all to all connectivity (each neuron receives connections from all neurons in a presynaptic population). The voltage below the spiking threshold for the excitatory neurons in the competing populations obeys (Eq. **S23**)

$$\frac{d}{dt}V(t)=-I_{syn}(t)-I_{adap}(t). \qquad \textbf{[S23]}$$

A neuron emits a spike when the voltage reaches the threshold $V_{th}=1$ (arbitrary units), after which the voltage is reset to $V_{reset}=0$. The model is endowed with a reflecting boundary at $V_{bound}=-1$ to avoid large negative excursions of the voltage. $I_{syn}(t)$ is the total synaptic current delivered to a neuron. $I_{adap}(t)$ is a slow adaptation current whose value is increased by $\Delta I_{adap}=0.1$ with each evoked spike and decays to zero exponentially with time constant $\tau_{adap}=1s$. Equations for the inhibitory populations are the same as that described above.

The synaptic currents to the excitatory (E) and inhibitory (I) subpopulations in population A are $I_{syn,E}(t)=I_{rec,A}(t)+I_{Inh,B}(t)+I_{ext,A}+I_{back}(t)$ and $I_{syn,I}(t)=I_{exc,A}(t)+I_{back}(t)$, respectively. Similar equations hold for population B. We assume that spikes generated in the network lead to δ-function currents in the postsynaptic neurons. The recurrent input generated by subpopulation A is $I_{rec,A}(t)=J_{EE}\sum_{i,j}\delta(t-t_i^j)$, where the index i runs over the neurons of subpopulation E of A, index j indicates spike timing, and $J_{EE}=0.012$. Inhibition into subpopulation E of A is generated by the I subpopulation of B, leading to $I_{Inh,B}(t)=J_{EI}\sum_{i,j}\delta(t-t_i^j)$, where now i runs over the neurons of subpopulation I of B with $J_{EI}=-0.07$. This (strong) inhibition is central to create the bistability in the network. Subpopulation I of A receives excitatory drive from the subpopulation E and $I_{exc,A}(t)=J_{IE}\sum_{i,j}\delta(t-t_i^j)$ with $J_{IE}=0.05$.

External inputs to the E subpopulations come from the relay population and a constant external background (Fig. 5*A*, rate-based model and Fig. S3*A*), and they are modeled as constant excitatory currents (Eqs. **S24** and **S25**)

$$\begin{array}{l} I_{ext,A}=I_0+(1-\varepsilon)(I_\lambda+I_v)+\varepsilon C(I_\lambda+I_v)^3 \\ I_{ext,B}=I_0-(1-\varepsilon)(I_\lambda+I_v)-\varepsilon C(I_\lambda+I_v)^3 \end{array}, \qquad \textbf{[S24 and S25]}$$

respectively, for populations A and B. As in the rate-based model, the parameter ε specifies the degree of nonlinearity in the inputs currents (in Fig. S3, $\varepsilon=0$ was chosen for the linear relay population and $\varepsilon=1$ was chosen for the nonlinear relay population), and $C=1/3s^2$ is a constant ensuring that the two terms in the sum have roughly the same magnitude for the values of the currents used in the stimulations (range between 0 and 1.5 s^{-1}). The choice of the cubic nonlinearity is consistent with the experimentally observed input to rate transfer functions in primary visual cortex (7). However, it should be noted that the particular choice of a cubic polynomial is not critical for the results shown in Fig. S3; qualitatively similar results were obtained with power between two and four. The value of the background current was $I_0=10s^{-1}$. The connectivity pattern of the relay population to the network was identical to that in the rate-based network.

Each E neuron received an independent source of noisy current modeled as Gaussian white noise $I_{back}(t)=\sigma\eta(t)$, where $\eta(t)$ is white noise with unit variance and $\sigma=3s^{-1/2}$. Each I neuron also received a negative mean current such that $I_{back}(t)=-\mu+\sigma\eta(t)$ with $\mu=10s^{-1}$.

The spiking network, as well as the rate-based network, is in the so-called noise-dominated regimen (10), where noise is the cause of the alternations (i.e., when the noise is completely removed from the network, the system does not oscillate and remains trapped in one of the attractors because of the weak adaptation current present in the model).

Numerical Procedures for the Neural Network Models. The dynamical equations for the energy and neural network models were integrated using Euler's method with time step $\delta t = 0.1$ ms. A shorter integration time step did not produce appreciable differences in any of the results that we obtained. The dominance durations for each percept in the energy model are defined by the amount of time in which the variable r is below (or above) $r = 0$. For the neural network model, a transition occurs when the firing rate of one population becomes larger (or smaller) than the firing rate of the other population. The energy, rate-based, and spiking neuronal models are typically run for 2,000 s (model time), which is close to the time run by the subjects in the experiments per condition. Means in all of the plots are computed from the time series generated with these long simulations, and error bars correspond to the SEM. We used custom C code to simulate the models (including a random generator for white noise that generated long nonrepetitive series) and Matlab to analyze and plot the data.

Sampling and Optimal Cue Combination for the Neural Network Models. In this section, we show that the neural network models with linear inputs sample the probability distribution defined in the inputs with a pure power law. In the text and *SI Methods, Attractor Neural Networks, Energy model*, we have shown that, in the energy model, the fraction of dominance of one percept is a sigmoidal function of the sum of the currents arising from independent cues (Eq. **3**) (Eq. **S26**)

$$f(s=A)=\frac{1}{1+e^{-2(I_1+I_2)/\sigma_{eff}^2}}, \qquad \textbf{[S26]}$$

where σ_{eff}^2 is the effective noise generated by the network and I_1 and I_2 are the currents corresponding to cues 1 and 2 (e.g., wavelength and speed cues in the previous sections). Through numerical simulations of the neuronal network model, we have shown that the fraction of dominance is also a sigmoid function of the sum of the currents (Fig. 6*B*). Therefore, Eq. **S26** is valid in general for attractor networks (i.e., noise-driven networks with attractor states), where σ_{eff}^2 depends on the details of the network.

To determine whether attractor networks sample from the posterior probability distribution over s encoded by the input current, we need to first specify how the input currents encode probability distributions. We show next that, when the inputs are probabilistic population codes (a particular type of neural code for encoding probability distributions), the posterior distribution over s given the input current is also a sigmoid of the input current, in which case the fraction of dominance $f(s=A)$ is indeed proportional to a power of the posterior distribution over s. In other words, the network would effectively sample the posterior distribution (raised to some arbitrary power).

To show this, we assume that for each cue i ($i = 1{,}2$), there are two presynaptic neural populations, one preferring stimulus interpretation A and the other preferring B. Thus, there are a total of four presynaptic populations. The presynaptic population A associated with cue i sends excitatory connections to the relay population (Fig. 5*A* and Fig. S3*A*) that feeds the attractor neural network while sending inhibitory connections to the relay population. The presynaptic population B has the reversed connectivity pattern.

In this section, we use the concept of probabilistic population codes (11, 12) (PPCs) to derive the posterior distribution over depth ordering given the presynaptic input, denoted $p(s|\vec{r}^{\,1}_A,\vec{r}^{\,1}_B,\vec{r}^{\,2}_A,\vec{r}^{\,2}_B)$, where $s = \{A,B\}$ are the possible percepts and $\vec{r}^{\,i}_A,\vec{r}^{\,i}_B$ are the firing responses of input populations A and B selective to cues $i = \{1{,}2\}$. The information present in the populations is combined optimally only if the posterior probability density function (pdf) over the stimulus parameters for each population i satisfies (assuming an uniform prior over s) (**S27**)

$$p(s|\vec{r}^{\,1}_A,\vec{r}^{\,1}_B,\vec{r}^{\,2}_A,\vec{r}^{\,2}_B)\propto p(s|\vec{r}^{\,1}_A,\vec{r}^{\,1}_B)p(s|\vec{r}^{\,2}_A,\vec{r}^{\,2}_B). \qquad \textbf{[S27]}$$

The distributions $p(s|\vec{r}^{\,i}_A,\vec{r}^{\,i}_B)$ can be expressed as (**S28**)

$$p(s|\vec{r}^{\,i}_A,\vec{r}^{\,i}_B)\propto p(\vec{r}^{\,i}_A,\vec{r}^{\,i}_B|s)p(s), \qquad \textbf{[S28]}$$

where $p(\vec{r}^{\,i}_A,\vec{r}^{\,i}_B|s)$ is the pdf that population i generates the activity patterns $\vec{r}^{\,i}_A,\vec{r}^{\,i}_B$ given that the stimulus is s and $p(s)$ is the a priori pdf over s and c_i. The a priori distribution is taken to be independent of s (that is, uniform in s). A uniform prior over s is indeed consistent with the fact that our subjects did not favor any particular depth ordering in our experimental results.

We assume that $p(\vec{r}^{\,i}_A,\vec{r}^{\,i}_B|s)$ belongs to the family of Poisson-like pdfs with sufficient linear statistics (Eq. **S29**),

$$p(\vec{r}^{\,i}_A,\vec{r}^{\,i}_B|s)=\Psi(\vec{r}^{\,i}_A,\vec{r}^{\,i}_B)\exp\left(\vec{h}_A(s)\cdot\vec{r}_A+\vec{h}_B(s)\cdot\vec{r}_B\right), \qquad \textbf{[S29]}$$

where Ψ is an arbitrary function that does not depend on s. This assumption is known to provide a good approximation to neural variability in vivo (11). We also assume that the kernels in s obey the symmetry condition $\Delta\vec{h}=\vec{h}_A(A)-\vec{h}_A(B)=\vec{h}_B(B)-\vec{h}_B(A)$.

Finally, it is easy to show using expressions **S27** and **S28** and Eq. **S29** that the posterior distribution over s is a sigmoid function of the presynaptic patterns of activity (Eq. **S30**)

$$p(s=A|\vec{r}^{\,1}_A,\vec{r}^{\,1}_B,\vec{r}^{\,2}_A,\vec{r}^{\,2}_B)=\frac{1}{1+e^{-\Delta\vec{h}\cdot(\vec{r}_A-\vec{r}_B)}}, \qquad \textbf{[S30]}$$

where we have defined $\vec{r}_s=\vec{r}^{\,1}_s+\vec{r}^{\,2}_s$. Note, first, that the probability over s depends only on the difference of the activities of populations A and B summed over the available cues and second, that it does not depend explicitly on the values taken by the cues.

Comparing Eqs. **S26** and **S30**, it is clear that the fraction of dominance generated by an attractor network can be equal to the posterior distribution over s as long as we equate the input current in Eq. **S26** to the proper combination of the probabilistic population codes $\vec{r}^{\,1}_A,\vec{r}^{\,1}_B,\vec{r}^{\,2}_A,\vec{r}^{\,2}_B$. Thus, if we set $I_1+I_2=w\Delta\vec{h}\cdot(\vec{r}_A-\vec{r}_B)$, where w measures the strength of the feed-forward connections, and use Eq. **S30**, then Eq. **S26** becomes (Eq. **S31**)

$$\begin{aligned}f(s=A)&=\frac{1}{1+e^{-2w\Delta\vec{h}\cdot(\vec{r}_A-\vec{r}_B)/\sigma_{eff}^2}}\\&\propto\left[p(s=A|\vec{r}^{\,1}_A,\vec{r}^{\,1}_B,\vec{r}^{\,2}_A,\vec{r}^{\,2}_B)\right]^{\frac{2w}{\sigma_{eff}^2}}.\end{aligned} \qquad \textbf{[S31]}$$

Therefore, the attractor network generates a dynamic that is indistinguishable from a sampling process of the probability distribution defined in Eq. **S30** raised to a power. By appropriately setting the value of w, it is possible to obtain any desired tempered sampling of the probability distribution. Note that this result is true only under the condition that the variability of neuronal activity lies in the exponential family with linear sufficient statistics (Eq. **S29**) as closely observed experimentally.

1. Moreno-Bote R, Shpiro A, Rinzel J, Rubin N (2008) Bi-stable depth ordering of superimposed moving gratings. *J Vis* 8:20.1–20.13.
2. Stoner GR, Albright TD, Ramachandran VS (1990) Transparency and coherence in human motion perception. *Nature* 344:153–155.
3. Hupé JM, Rubin N (2003) The dynamics of bi-stable alternation in ambiguous motion displays: A fresh look at plaids. *Vision Res* 43:531–548.
4. Bishop CM (2006) *Pattern Recognition and Machine Learning* (Springer, Berlin).
5. Risken H (1989) *The Fokker-Planck Equation* (Springer, Berlin), 2nd Ed.
6. Kramers HA (1940) Brownian motion in a field of force and the diffusion model of chemical reactions. *Physica* 7:284–304.
7. Priebe NJ, Mechler F, Carandini M, Ferster D (2004) The contribution of spike threshold to the dichotomy of cortical simple and complex cells. *Nat Neurosci* 7:1113–1122.
8. Markram H, Tsodyks M (1996) Redistribution of synaptic efficacy between neocortical pyramidal neurons. *Nature* 382:807–810.
9. Abbott LF, Varela JA, Sen K, Nelson SB (1997) Synaptic depression and cortical gain control. *Science* 275:220–224.
10. Moreno-Bote R, Rinzel J, Rubin N (2007) Noise-induced alternations in an attractor network model of perceptual bistability. *J Neurophysiol* 98:1125–1139.
11. Ma WJ, Beck JM, Latham PE, Pouget A (2006) Bayesian inference with probabilistic population codes. *Nat Neurosci* 9:1432–1438.
12. Beck JM, et al. (2008) Probabilistic population codes for Bayesian decision making. *Neuron* 60:1142–1152.

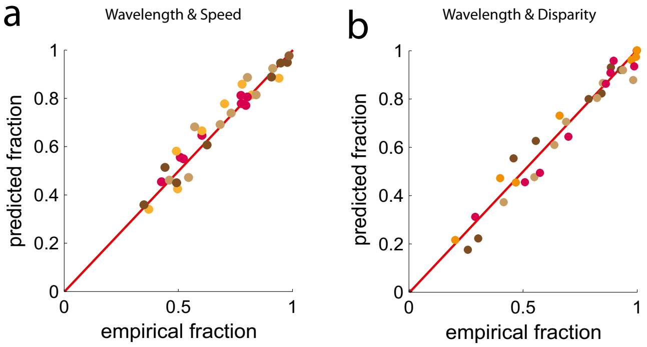

Fig. S1. (*A*) Experimental vs. predicted fractions of dominance when using wavelength and speed. Each color corresponds to one subject. (*B*) Experimental vs. predicted fractions of dominance when using wavelength and disparity.

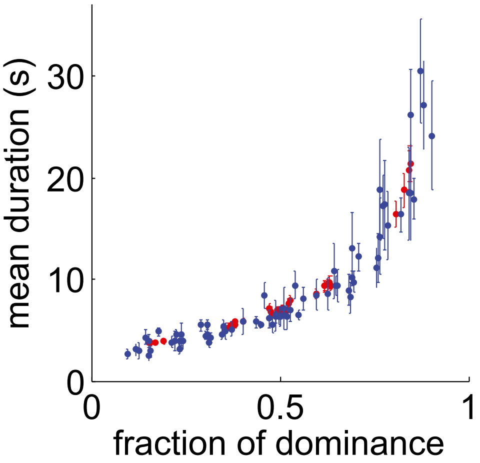

Fig. S2. Mean dominance duration of a percept as a function of the fraction of dominance of the percept for the experimental data averaged across subjects (blue) and for the model with nonlinear activation function (red).

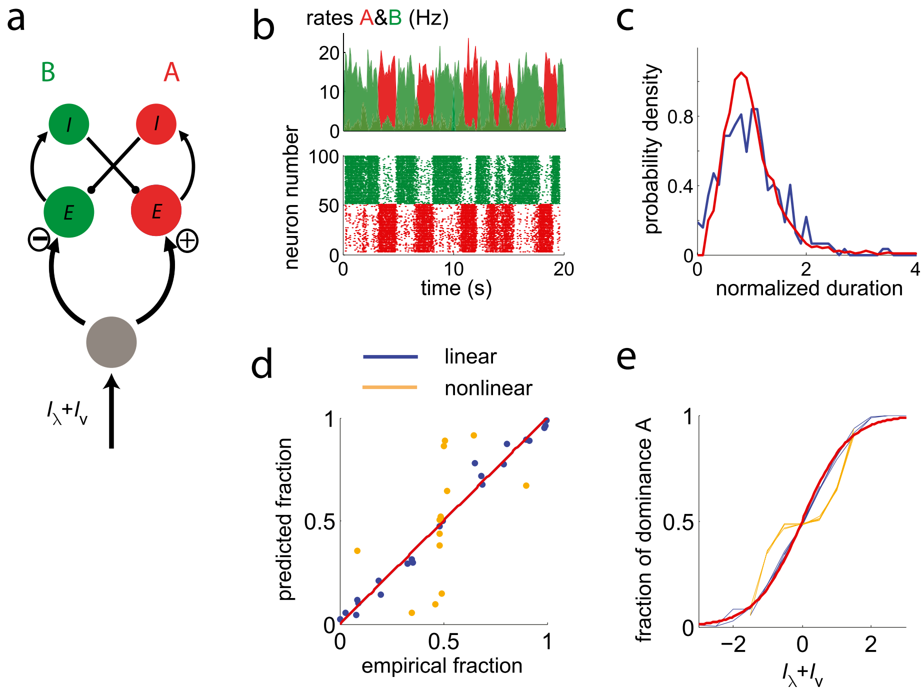

Fig. S3. Sampling in a spiking attractor network. (*A*) Architecture of the network. Two excitatory populations that receive inputs from a relay population compete for dominance through mutual inhibition mediated by local inhibitory networks. (*B*) Excitatory population firing rates as a function of time for the neural network model (*Upper*) and raster plot including all neurons in the network (*Lower*). Red, neurons encoding percept *A*; green, neurons encoding percept *B*. (*C*) The distribution of dominance durations from the model in the neutral condition (red) and from the pooled data across subjects (blue) for the wavelength-speed experiment in the neutral condition ($n = 320$). Time has been normalized so that the mean of the distributions is one. Both distributions are close to each other and have a coefficient of variations close to 0.6. Because the distribution from the model corresponds to the case in which the cues are neutral, it is the same regardless of whether the activation function of the relay unit is linear or nonlinear. (*D*) Predicted fractions using the multiplicative rule vs. empirical fractions of dominance generated by the neural network with linear (blue dots) and nonlinear (orange dots) relay population. (*E*) The fractions of dominance are well-approximated by a sigmoid function of the sum of input currents to the relay population when the relay population is linear (blue) but not when it is nonlinear (orange). Red curve corresponds to the best sigmoid fit in the linear case.

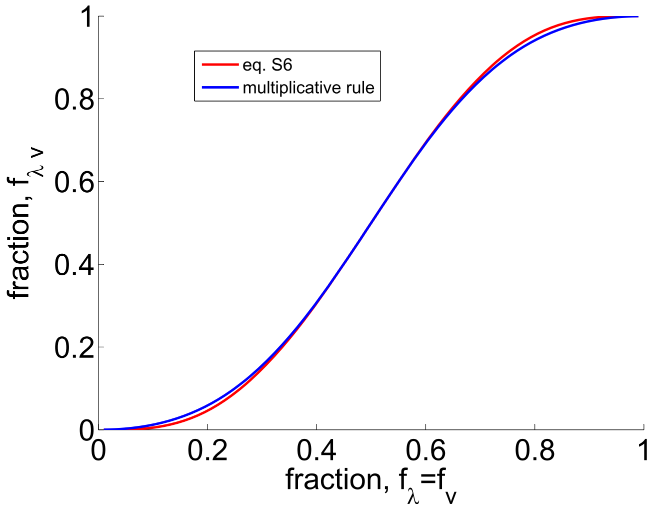

Fig. S4. Comparison between the multiplicative rule (Eq. **1**) and its continuous version (Eq. **S6**), showing that the two expressions lead to nearly identical predictions. In both cases, $f_{\lambda v}$ is plotted as a function of $f = f_\lambda = f_v$.