Published in final edited form as:
J Neurophysiol. 2007 September ; 98(3): 1125–1139. doi:10.1152/jn.00116.2007.

Noise-Induced Alternations in an Attractor Network Model of Perceptual Bistability

Rubén Moreno-Bote[1], **John Rinzel**[1,2], and **Nava Rubin**[1]

1*Center for Neural Science, New York University, New York*

2*Courant Institute of Mathematical Sciences, New York University, New York*

Abstract

When a stimulus supports two distinct interpretations, perception alternates in an irregular manner between them. What causes the bistable perceptual switches remains an open question. Most existing models assume that switches arise from a slow fatiguing process, such as adaptation or synaptic depression. We develop a new, attractor-based framework in which alternations are induced by noise and are absent without it. Our model goes beyond previous energy-based conceptualizations of perceptual bistability by constructing a neurally plausible attractor model that is implemented in both firing rate mean-field and spiking cell-based networks. The model accounts for known properties of bistable perceptual phenomena, most notably the increase in alternation rate with stimulation strength observed in binocular rivalry. Furthermore, it makes a novel prediction about the effect of changing stimulus strength on the activity levels of the dominant and suppressed neural populations, a prediction that could be tested with functional MRI or electrophysiological recordings. The neural architecture derived from the energy-based model readily generalizes to several competing populations, providing a natural extension for multistability phenomena.

Introduction

When observers are presented with an ambiguous stimulus that has two distinct interpretations, their perception alternates over time between the different possible percepts in an irregular manner, a phenomenon known as *perceptual bistability*. Bistability arises in many domains of perception: ambiguous figures (Necker 1832), figure–ground segregation (Rubin 1921), ambiguous motion displays (Hupé and Rubin 2003), auditory segmentation (Pressnitzer and Hupé 2006), and—the domain that has been studied most extensively—binocular rivalry (Blake 1989, 2001; Levelt 1968; Logothetis 1998; Tong 2001; Wheatstone 1838). In addition to generating much experimental work, binocular rivalry has attracted much attention theoretically, and many models have been proposed for it (Bialek and DeWeese 1995; Blake 1989; Dayan 1998; Freeman 2005; Laing and Chow 2002; Lehky 1988; Lumer 1998; Wilson 2003). Although there are fewer quantitative studies of other bistable perceptual phenomena, there is evidence that they share many properties of binocular rivalry alternations (Rubin and Hupé 2004; van Ee 2005). Thus models of binocular rivalry may be generalized to other bistable perceptual phenomena.

Although the alternations seem haphazard, for a fixed stimulus the durations are drawn from a stationary distribution that resembles a skewed Gaussian, typically fit by a gamma or log-normal function (e.g., Lehky 1995; Levelt 1968; Rubin and Hupé 2004). Importantly,

Address for reprint requests and other correspondence: R. Moreno-Bote, Center for Neural Science, New York University, 4 Washington Place, Room 1102, New York, NY 10003 (E-mail: E-mail: rmoreno@cns.nyu.edu).

alternations occur not only when the two percepts are balanced in strength (equal mean dominance durations), but also when one is significantly stronger than the other. When stimulus parameters that affect the relative strength of the two interpretations are varied continuously, the relative time spent perceiving each changes gradually (Hupé and Rubin 2003; Levelt 1968). In the domain of binocular rivalry, where the strength of each competing percept can be manipulated independently (e.g., by the contrast of the monocular images), two additional important observations were summarized by Levelt (1968). His “Proposition II” states that the imbalance in dominance time caused by weakening only one image (while keeping the other fixed) occurs mainly through an increase of the mean dominance duration of the other (unchanged) image, with little or no effect on the dominance durations of the manipulated one. (Note that this is a statement about absolute mean durations; the fractions of dominance time of each percept obviously both change because they must add up to one.) Levelt's “Proposition IV” states that when the monocular images are strengthened simultaneously, the mean durations of both eyes decrease (i.e., the rate of alternations increases, although the fraction of time spent perceiving each image remains unchanged).

What causes the alternations? Although this is a central question about perceptual bistability, the mechanisms underlying the perceptual switches are not well understood. In most current models alternations between dominance of two or more competing neuronal populations arise from some form of slow adaptation acting on the dominant population, either in its firing rate or in its synaptic output (synaptic depression) or both, which leads to a switch in dominance to the competing population (Kalarickal and Marshall 2000; Lago-Fernandez and Deco 2002; Laing and Chow 2002; Lehky 1988; Matsuoka 1984; Stollenwerk and Bode 2003; Wilson 2003). In the absence of noise or finite-sized induced fluctuations, models in which switches are caused by adaptation generate alternations with perfect periodicity; we therefore term them *oscillator models*. Importantly, in such models noise is assumed to be an inessential (albeit experimentally inevitable) component of the perceptual alternations.

An alternative possibility is that the main cause of perceptual switching is noise—external, internal, or both. Noise, in the form of unavoidable perturbations, is ubiquitous in the brain at multiple scales, from vesicular release and spiking variability to fluctuations in global neurotransmitter levels. Furthermore, external noise can cause perceptual alternations (cf. Kanai et al. 2005; Lankheet 2006) as can some internally generated noise (e.g., blinks). This raises the possibility that noise is the primary cause for alternations. In this scheme, dominance of each of the competing percepts can be viewed as a stable state of the neuronal dynamics (i.e., attractor; Hertz et al. 1991), with noise causing the system to alternate between them. We therefore term these *noise-driven attractor models*. (Additional involvement of neural adaptation may still be present, but in this scheme it would not play the primary role and would not lead to alternations in the absence of noise.) Attractor models make a fundamentally different prediction than do oscillator models about the consequence of eliminating noise from the system: rather than showing perfectly periodic alternations, they predict that the perceptual alternations would cease—i.e., the system would settle down in one of the two percepts and stay there indefinitely. Although this is a thought experiment that cannot be performed practically, exploring the distinction between the two alternatives theoretically is important for our understanding of the underlying mechanisms.

Herein we present an attractor-based framework in which alternations are induced by noise and are absent without it. The proposal that bistable transitions may be mediated by noise has been made before (e.g., Brascamp et al. 2006; Freeman 2005; Haken 1994; Kim et al. 2006; Lankheet 2006; Riani and Simonotto 1994; Salinas 2003). Our work goes beyond previous models of noise-driven bistability by constructing a neurally plausible attractor model that produces behaviors consistent with the experimental findings summarized earlier. A particular challenge is posed by Proposition II introduced by Levelt (1968) because it implies that

increasing input strength to one attractor reduces the energy barrier for the other attractor (Kim et al. 2006), and such behavior does not arise in commonly used energy functions (see, e.g., Hertz et al. 1991). We therefore start by formulating a simple two-well energy function that includes coupling between the input strength to one attractor and the energy barrier of the other. We then derive from the energy function dynamical equations of a rate-based (mean-field) model. The equations suggest a novel network architecture, where information about the input strength of each percept is sent not only to the population representing it but also to the population representing the competing percept. This, in turn, leads to the novel prediction that increasing stimulus strength to one population will reduce the activity level of the competing population by recruiting more inhibition while it is dominant. Finally, we show that the model can also be realized in a spiking neuronal attractor network, using the neuronal architecture derived for the rate-based model, thus providing a more realistic description of the neuronal dynamics during perceptual bistability.

Results

We start by positing that each of two neuronal populations, labeled *A* and *B*, represent a different possible interpretation of the stimulus. The neural correlate of competition for perceptual dominance is a competition between these populations for higher activity. The activities of the populations are described by their mean firing rates, r_A and r_B. We denote by A_{on} and B_{on} the states of dominance of populations A ($r_A \gg r_B$) and B ($r_B \gg r_A$), respectively.

Hypothesizing noise-driven alternations means a fundamentally different structure of the trajectories in state space (the space of neuronal activities). This is illustrated in the *left* and *right panels* in Fig. 1A, which visualize on the plane of population firing rates (r_A, r_B) the evolution of the two models over time. (Time is not explicit in this representation: rather, points along the trajectory correspond to snapshots of the state of the system at regular time intervals.) In both oscillator and noise-driven attractor models, perceptual alternations correspond to alternations between points A_{on} and B_{on}. However, the trajectories in state space that move the system between these states are fundamentally different in the two cases. In oscillator models (Fig. 1A, *left*), the alternations follow a cyclic trajectory in the plane (r_A, r_B), caused by the deterministic effect of the slow negative feedback provided by the adaptation. Thus the system is characterized by a limit cycle, with a large proportion of the time spent around A_{on} and B_{on}. In contrast, in our noise-driven model (Fig. 1A, *right*) A_{on} and B_{on} are stable fixed points, or attractors, like those obtained from a stimulus that activates only one of the populations (e.g., by turning off one of the monocular images in binocular rivalry). All trajectories in state space approach either A_{on} or B_{on}, depending on which side of the diagonal (separatrix) they originate from. The alternations between the two stable states are attributed to noise that occasionally allows the system to overcome the energy barrier between them. In the presence of noise, dominance alternates over time as the system visits the states A_{on} and B_{on} in turn (Fig. 1, *center plot*, oriented diagonally). In the absence of noise, the system would flow into one of the stable states (depending on initial conditions) and stay there indefinitely.

Because bistable perceptual alternations are not regular, but rather appear stochastic, many oscillatory models also assume a role for noise (Kalarickal and Marshall 2000; Lago-Fernandez and Deco 2002; Lehky 1988; another proposal is that the irregularity of alternations arises from finite-size effects; Laing and Chow 2002). It is therefore important to sharpen and clarify the distinction we make between two types of models, and the two panels in Fig. 1A are particularly useful for that.

We use the term “noise-driven attractor model” to refer to any system where the points A_{on} and B_{on} in state space are stable fixed points, i.e., a system that will not undergo alternations between these states in the absence of noise. Such a system may or may not also contain

adaptation, as long as the adaptation is not strong enough to drive alternations by itself (i.e., when noise is eliminated). Indeed, as will be seen later, in our model we use weak adaptation to adjust the form of the distributions of dominance durations so that they resemble those observed experimentally. However, because this adaptation is too weak to drive alternations when noise is eliminated, its inclusion does not change the noise-driven nature of the model.

Conversely, by "oscillatory model" we refer to one where the points A_{on} and B_{on} are *not* stable fixed points, but rather belong to a limit cycle that is the only stable state (when both populations A and B are stimulated). Such a system may or may not also contain noise (e.g., to introduce jitter in the dominance durations) so long as the noise does not destroy the stable limit cycle. (One may construct more complex systems where state space changes over time from having attractors to oscillatory stable states; such systems would not fall into either preceding category and we do not study them here.)

A one-variable energy model for bistability

The two-attractor structure of state space proposed in Fig. 1A (*right*) naturally leads to a description by an energy function with two local minima, corresponding to A_{on} and B_{on}, and a barrier corresponding to the separatrix (Fig. 1B). We therefore first sought to find an energy-based formalism to describe the dynamics of perceptual alternations. This formulation will later shed light on how to build more realistic rate-based and spiking neural networks. The observations summarized by Levelt (1968) about the effect of stimulation strengths on the mean dominance durations of each percept present an important challenge in the construction of an energy function. A simplistic extension of commonly used energy functions would produce a system where increasing the input strength to percept A would deepen its own minimum. This would make it harder for the system to escape from A, which in turn would increase its mean dominance durations. This is at odds with Levelt's Proposition II, which states that the main effect of increasing the input strength to A is a decrease in the mean dominance duration of B. As observed by Kim et al. (2006), the latter behavior implies that an increase of the input to A has the effect of heightening the energy barrier of the population representing percept B (Fig. 2A, *right*). Similarly, Proposition IV (alternation rate grows as both stimuli are strengthened) implies that increasing both inputs lowers—not heightens—the energy barrier (Fig. 2A, *left*). A simple energy function that has these two properties is

$$E(\Delta r) = \Delta r^2\left(\Delta r^2 - 2\right) + g_A(\Delta r + 1)^2 + g_B(\Delta r + 1)^2 \quad (1)$$

where the single variable $\Delta r = r_A - r_B$ is the difference between the firing rates of the two competing populations and g_A and g_B are their input strengths. The minima are located close to $\Delta r = \pm 1$ (states A_{on} and B_{on}, respectively; for simplicity, the firing rates are dimensionless here). The first term of the energy function ensures that there are two local minima for small values of the stimulus strengths. The next two quadratic terms are proportional to the stimulation strengths; each increases the energy of the competing minimum without changing its own minimum energy.

For a model based on an energy function the dynamic variable satisfies $d\Delta r/dt = -\tau^{-1} dE(\Delta r)/d\Delta r$. This means that the dependent variable Δr moves along the horizontal axis of the energy function (Fig. 2A) toward the location of the closest minimum with a velocity proportional to the slope of the function. Because the slope of the energy function is zero at the minima, those points are fixed points of the dynamics. In addition to the deterministic rule specified earlier, we introduce a noise source to allow random transitions between the minima. The time evolution is therefore given by

$$\tau \frac{\mathrm{d}}{\mathrm{d}t}(\Delta r) = -4\Delta r\left(\Delta r^2 - 1\right) - 2g_A(\Delta r - 1) - 2g_B(\Delta r + 1) + n(t) \quad (2)$$

Here τ (set to 10 ms in the subsequent simulations) is the timescale in which Δ*r* changes and *n*(*t*) is a colored Gaussian noise (see Eq. A1, APPENDIX A). Because perception of stimulus *A* happens whenever the firing rate of population A is higher than that for population B ($r_A > r_B$), an alternation occurs when the variable Δ*r* crosses zero. This dynamics generates trajectories that linger for a short while around one of the fixed points and then move to the other (Fig. 1B, *middle*, oriented diagonally).

When alternations between two states are driven purely by noise, the lifetimes of each state are distributed exponentially (Kramers 1940). With refractory period or other biophysical constraints, the distribution of dominance durations is nearly exponential, with the peak determined by the timescale of the noise (e.g., Kramers 1940; van Kampen 2001). As can be seen in Fig. 1C, this is also the case for the model as formulated in Eq. 2, where the peak of the distribution of dominance durations is at the timescale of the noise we used, approximately 100 ms. This is very far from experimentally observed distributions, which typically peak at timescales of seconds and have a shape resembling a skewed Gaussian (Lehky 1995; Levelt 1968). One may propose to address this by assuming that perceptual alternations are driven by noise sources that act at a slower timescale, of seconds (e.g., originating from endogenous attention modulations and/or global neurotransmitter levels). This approach is limited, however: while it can certainly lengthen the dominance durations, it is not sufficient to fit the shape of their distributions (see following text, Fig. 7). We therefore favor another approach, which can yield both realistic means and shape of distributions of dominance durations. We propose that biophysical noise sources characterized by fast timescales (~100 ms) do play a major role in causing alternations. However, unlike in the simple model of Eq. 2, we suggest that in reality there are additional mechanisms that effectively reduce the probability that the system leaves an attractor right after it has settled into it, compared with the probability of later transitions. There is independent evidence for the existence of such “short-term persistence” mechanisms (see DISCUSSION and Leopold et al. 2002), but their precise nature is not well understood. In the model presented in the following sections, we achieve this tendency by adding a weak adaptation current. The initial activity level of the dominant population (i.e., immediately after transitions) will be too high for the noise to push the system to the competing state. Over time, however, the weak adaptation will bring the activity to a slightly lower level, comparable to that of the noise amplitude, so that the probability of transition will increase. Importantly, however, in our model adaptation alone (i.e., without noise) will not be enough to cause alternations, i.e., they will still be noise driven. In terms of the state-space and the energy landscape (Fig. 2), the effect of adaptation will be to add a slow, time-dependent forcing that effectively reduces the depth of the minimum associated with the dominant percept over time. However, the adaptation will be too weak to destroy the energy minima (i.e., to destabilize the states A_{on} and B_{on}), and therefore noise will still be crucial for dominance switches.

The mean dominance durations of each attractor state, calculated from simulations of Eq. 2 for different input strengths, show that the system indeed satisfies Levelt's propositions (Fig. 2B, solid lines). This is a direct consequence of our choice of energy function, specifically the dependence of the height of energy barriers on input strength. This dependence arises from the two terms where the input strengths (g_A and g_B) are multiplied with the state variable (Δ*r*). Although the effect of g_B on T_A is much larger than that on T_B (Fig. 2B, *right*), the effect on T_B is not negligible. This is because, although increasing g_B greatly reduces the energy barrier for A_{on}, it also slightly increases the barrier for B_{on}. This behavior is consistent with experimental results (Brascamp et al. 2006). In the next section we will see that the coupling

between the input strength to one population and the energy barrier of the other, posited to obtain the experimentally observed dependencies of mean dominance durations on stimulus strength, motivates a novel network architecture and leads to novel predictions about the levels of activity of the neural populations.

Derivation of a rate-based model and network architecture

In this section we construct a rate-based network model based on the energy description of Eq. 1. We first extend Eq. 1 to a two-variable energy function (Eq. B1, APPENDIX B). This energy describes the dynamics of two populations, *A* and *B*, through their firing rates r_A and r_B. We then derive from the two-variable energy function two coupled differential equations describing the dynamics of the two populations' firing rates in the presence of noise (Eq. B4).

The dynamics equations determine the time evolution of the firing rates of the two populations and can be interpreted as originating from an underlying neural network. Indeed, the neural populations in the architecture presented in Fig. 3A obey the dynamics derived from the two-population energy function. Each population has recurrent excitation and each inhibits the other through direct cross-connections. (Although the schematic indicates that both excitation and inhibition emanate from a single population, this connectivity could be achieved with excitatory and inhibitory subpopulations; not shown.) The network shares a basic feature with many other models of bistability: to ensure that only one population is active at any time ("mutual exclusivity"; Leopold and Logothetis 1999;Rubin 2003), mutual inhibition is exerted between the two populations (Blake 1989;Laing and Chow 2002;Wilson 2003). Our model, differing from some others, requires strong recurrent excitatory connections to produce robust winner-take-all behavior for relatively weak inputs. However, for very weak inputs a single low-activity resting state is the only attractor.

A novel feature of the model that is clearly visible in the architecture is that the local inhibitory subpopulations (small circles in Fig. 3A) are driven by the total external stimulation. A crucial point here is that the external input to these subpopulations constitutes not only a copy of the external input to "their" excitatory population, but also the input sent to the competing population (e.g., to the other eye). Moreover, the rate-based equations (Eq. B4) require that this total external input be gated by the activity level of the corresponding excitatory population, so that each recurrent population *k* (*k* = *A*, *B*) receives back inhibition equaling $(g_A + g_B)r_k$. This feature is a consequence of the multiplicative terms $g_k(\Delta r \pm 1)^2$ in the one-variable energy function (Eq. 1). Recall that those terms were required to make the model behave in accordance with propositions II and IV of Levelt (1968). Inspection of Eq. B4 now sheds light on how the multiplicative terms give rise to these behaviors. Increasing the input to one population, say A, results in stronger inhibition to it when it is dominant (i.e., when $r_A = 1$) and also in stronger inhibition to population B when the latter is dominant (i.e., when $r_B = 1$). At the same time, the increase of g_A also provides additional excitatory input to population A, and therefore the total input to it remains largely unaffected when it is dominant. In contrast, population B does not enjoy stronger excitatory input from the increase in g_A, and therefore its total input, although dominant, is reduced by an amount $-g_A r_B$. Consequently, the mean dominance duration of B is reduced because less noise is now required to kick it out of dominance; meanwhile, the mean dominance duration of population A remains nearly unchanged (Levelt's proposition II) because there is not much change to its total input while dominant. This argument does not apply when the input to A is so large that state B_{on} is close to disappearing and A_{on} becomes the only stable state of the system (see last subsection in RESULTS). Similarly, simultaneous increases of the input strength to both populations cause enhanced inhibition to both during dominance, and therefore an increase in alternations rate (Levelt's proposition IV). As for the question how the multiplicative terms $(g_A + g_B)r_k$ may be implemented, they can be realized in a biophysically plausible way by a nonlinear input–output transfer function for the neurons

of the inhibitory subpopulations (see, e.g., the quadratic function in the next version of the model, Eqs. B5–B7).

Finally, we modify the architecture to achieve more plausible generalization to perceptual multistability, i.e., when the number of competing percepts N is >2 (Rubin 2003; Suzuki and Grabowecky 2002). A simplistic generalization of the architecture in Fig. 3A would require each of the N populations to send direct inhibitory connections to all other populations, causing the number of connections to grow as N^2 and implying that each population needs to have knowledge of all its potential competitors. These problems are solved by the alternative architecture shown in Fig. 3B, which consists of a common neural pool that is driven by all of the external inputs, and sends by excitatory connections information about the total summed input to all of the local inhibitory subpopulations, which in turn inhibit their respective excitatory populations as discussed earlier. This eliminates the need for direct connections between the neural populations representing the different percepts and reduces the number of required connections from $O(N^2)$ to $O(N)$.

Dynamics of the noise-driven rate-based model and the role of weak adaptation

We have simulated a two-population rate-based model using the architecture in Fig. 3B with the addition of weak adaptation currents (for details see APPENDIX B, second section). Figure 4 presents time courses for the relevant dynamical variables of an excitatory neuronal population that undergoes an alternation from the suppressed to the dominance state and back to the suppressed state. Traces are shown for two different conditions: weak (gray) and strong (black) stimulation. [Equal stimulation was applied to the two populations in each case; to facilitate direct comparison between the two conditions, we used the same noise $n(t)$ for both simulations.]

We first use Fig. 4 to further explain the effect of the weak adaptation in our model because it is fundamentally different from that in oscillatory models. The dashed curved traces in Fig. 4, A and B show the activity level of the dominant population and of the total input to it, respectively. A slight decline over time is clearly evident in the noise-free system, but it also exists for the mean activity and mean total input in the presence of noise. This decline is caused by the gradual increase of the adaptation current (Fig. 4D; the adaptation does not exhibit rapid fluctuations because it integrates the activity slowly; cf. Eq. B5). Because the total input to the excitatory population is given by

$$\text{Total Input} = \text{Recurrent Excitation} - \text{Inhibition} - \text{Adaptation Current} + \text{Stimulus} + \text{Noise}$$

(cf. Eq. B5), its mean decreases as the adaptation current increases over time. Note, however, that the asymptotic value of the total input is well above the transition threshold of the system (horizontal line in Fig. 4B). The adaptation is therefore not sufficient to drive dominance switches by itself. Instead, transitions occur by chance, when noise-evoked fluctuations bring the total input below the threshold. Thus if noise is removed from the model, the system would never show alternations. This noise-driven switching mechanism is fundamentally different from what happens in oscillator models: in those, the adaptation is taken to be strong enough to cause switching in dominance by reducing the total input below the transition threshold, even in the absence of noise.

Although the weak adaptation does not drive alternations in our model, it serves another important purpose: it provides a mechanism to make the probability of transition time dependent. Because the mean and the amplitude of the fluctuations in our model do not change

over time, without adaptation the probability that the noise would cause the total input to dip below threshold would have been constant in time. This, in turn, would have yielded exponential-like distributions of dominance durations, peaking at the timescale of the noise (recall that without weak adaptation brief dominance durations of ~100 ms are much more likely to occur; Fig. 1C). The weak adaptation provides a time-varying mean input that disfavors early transitions in comparison with later transitions, thus providing a form of "short-term persistence."

In terms of the energy landscape (Fig. 2), the weak adaptation can be thought of as causing slow changes in the shapes of the energy wells around A_{on} and B_{on} and the energy barrier between them (but without completely destabilizing the two attractors). Specifically, over time there is a decrease in size of the basin of attraction associated with the dominant percept, together with a shift of the separatrix (approximately the peak of the energy barrier) toward the same attractor. Figure 5 provides an example of an individual trajectory of the system, illustrating a transition from B_{on} to A_{on} on the plane (r_A, r_B), the change in the location of the seperatrix over time, and the absence of transition without noise.

Returning to Fig. 4, we next examine it to gain further understanding of the dependence of dominance durations on stimulation strength. During dominance, the activity level and the total input (Fig. 4, A and B, respectively) are both slightly higher in the weak stimulation condition (*gray traces*) than in the strong stimulation condition (*black traces*). This, in turn, results from a higher inhibition when the competing stimulus is stronger (Fig. 4C). As a result, the total input during dominance is closer to the threshold for stronger competing stimuli, so that transitions tend to occur sooner, in accordance with Levelt's propositions II and IV. The dotted lines in Fig. 2B confirm that the rate-based model indeed obeys these propositions over a wide range of stimulus strengths. For weak enough inputs, alternation behavior gives way to quiescence: the bistable attractor states disappear and a single (resting) low-activity attractor state for both populations is the only available firing pattern. The transition regime between the alternation mode (A_{on} or B_{on}) and quiescence is sharp, and it is characterized by the presence of random sequencing between three states: A_{on} alone, B_{on} alone, and the resting state. For weak inputs, the resting state dominates most of the time, whereas for less weak inputs the A_{on} and B_{on} states alternate. Although sharp, the transition regime is continuous, with the resting state occupying an increasing fraction of time as the stimulus strength decreases. One may interpret this continuous transition as corresponding to the stimulus detection threshold. For very large input strengths, the system can oscillate even without noise because of the presence of weak adaptation; for yet stronger stimuli, steady coactivity of the two populations occurs (the latter is also observed in adaptation-based models; Shpiro et al. 2007). However, such large inputs are not likely to be experienced in reality because of gain control mechanisms that operate at multiple levels of sensory processing. The existence of a large range in which a winner-take-all regime is present between the low- and high-input strength regimes is controlled by the strength of the recurrent connections of the excitatory populations. We have established a set of conditions for the network connectivity parameters that approximately determine when the attractor states exist (APPENDIX B, third section).

To further examine the effect of noise on dominance transitions we calculated averages of the time courses of input noise synchronized to specific transition events ["switch-trig-gered-averages" (STAs)]. The solid curve in Fig. 6A shows the STA for transitions from suppressed to dominant states of one population (arbitrarily chosen) and the dashed curve shows the STA for the reverse transitions of the same population. The curves indicate that transitions tend to occur when there are simultaneous increases in input noise to the population switching to dominance and decreases in input noise to the population becoming suppressed. (Note that the transitions occur with a short delay after the coincidental noise fluctuations in the two populations, reflecting the neuronal integration timescale.) Figure 6B shows that this tendency

holds for individual transitions, not just for averages. Each point in the figure represents the values of the input noise to population A against that of population B at moments of transition. In spite of the variability in the noise values at individual transition events, there is a clear and an almost complete separation between the two clouds, with the dot symbols, indicating transitions of population A from suppressed to dominant, clustering in the *bottom right quadrant* (i.e., when $n_A > 0$ and $n_B < 0$), and the cross symbols, indicating the opposite transitions, clustering in the opposite quadrant. Furthermore, the clouds of points are elongated with slope near one, suggesting that a stronger-than-average positive fluctuation in the input to A can push it to dominance even if B receives a weaker-than-average negative fluctuation, and vice versa.

Recently, Lankheet (2006) conducted an experiment to test the effect of modulations in the external (stimulus) noise on perceptual transitions. Two random-dot kinematograms with different directions of motion were used as competing stimuli in a binocular rivalry paradigm. The coherence levels were modulated in a pseudorandom fashion as observers continually indicated their percept. Lankheet then calculated the STAs of the coherence in the two stimuli. STAs associated with transitions from suppression to dominance of an eye showed a peak just before the transition and those of the other eye show a negative (if weaker) peak. The experimental STAs resemble our simulated STAs in Fig. 6A. At the same time, there are a few notable differences between Lankheet's results and the STAs shown in Fig. 6A. First, only one of Lankheet's subjects showed a negative peak in the STAs of the transitions from dominance to suppression, whereas the STAs produced by our model show positive and negative peaks of approximately the same height. Dissimilar heights can be obtained in a slightly modified version of our model, too, by injecting the noise directly into the inputs of all populations that receive external stimulation, rather than as a perturbation to the excitatory populations only (not shown), which is more similar to the Lankheet (2006) paradigm of perturbing the external stimuli. Second, some of Lankheet's subjects showed wide and shallow peaks in their STAs several seconds before the narrow peaks immediately preceding the transition, which are not observed in our simulations. To mimic the experimental observations, we ran a simulation with a noise timescale of 500 ms as that used in the experiment (instead of the much shorter 100-ms timescale used to compute the STAs in Fig. 6A). The new STAs resemble those found experimentally, including the presence of wide and shallow peaks preceding the sharp peaks right before transitions (Fig. 6C, solid lines). Using simulation results of a competition model, Lankheet (2006) interpreted the shallow peaks as the consequence of firing rate adaptation. However, even after we removed adaptation from our model altogether (and increased the noise level so that the alternation rate is kept constant, around 0.25 Hz), the shallow peaks did not disappear (Fig. 6C, dashed lines). This suggests that adaptation is not necessary to produce those peaks.

The interplay between noise and adaptation levels and its effect on the distribution of dominance durations

With appropriate choice of the amplitudes of noise and adaptation current, the rate-based Eq. 2 produces noise-driven switches whose distribution of dominance durations (Fig. 7A) agrees with those typically observed during rivalry, being well fit by gamma or log-normal functions (Lehky 1988;Levelt 1968). The timescale and amplitude of both noise and adaptation affect the shape of the distributions. Figure 7B presents the distributions obtained for two other conditions, stronger and weaker adaptation (dashed and dotted lines, respectively; conditions were compared with the mean dominance durations kept approximately constant, which means that as adaptation strength was increased, the noise amplitude was reduced accordingly.) When adaptation is strong, alternations are dominated by the dynamics of this outward current, making the durations less variable. The distribution becomes narrower and symmetrical around the mean dominance duration (Fig. 7B, thin solid curve). The limiting case of strengthening

adaptation (relative to noise amplitude) produces a noise-free oscillatory system; i.e., the distribution of dominance durations becomes a delta function (not shown). At the other extreme, when adaptation is removed altogether, the distribution becomes severely skewed with its peak shifted down to a value closer to the timescale of the noise (100 ms in our case). These results indicate that to obtain realistic distributions of dominance durations adaptation should be present, but weak. Importantly, the values of adaptation and noise in our model that produce realistic distributions are such that adaptation cannot generate switches by itself, i.e., when noise is removed from the system. In the presence of adaptation there should be correlations between the durations of consecutive percepts, but because in our model the adaptation is weak, correlations are small (not shown), in accordance with experimental evidence (Fox and Herrmann 1967;Rubin and Hupé 2004). Nevertheless, the role of adaptation is important and twofold. It produces a time-dependent probability of transitions that gives realistic distributions of dominance durations. Also, adaptation's slow time-scale in companion with the noise amplitude sets the timescale (seconds) of alternations.

Bistability and alternations in a spiking neural attractor network

In this section we present results from simulations of a cell-based network with spiking neurons based on the rate model presented above (see APPENDIX C). The architecture was like that in Fig. 3B (without population C), with 100 neurons per population. We used leaky integrate-and-fire neurons with weak adaptation currents (to obtain dominance durations consistent with experimental observations; see above). The connectivity between neurons in each stimulus-selective excitatory population was all to all. In addition, each neuron projected to all other neurons in a target population. Background synaptic conductance input was modeled using fast kinetics like those of α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid (AMPA) and γ-aminobutyric acid type A ($GABA_A$) receptors receiving white noise, uncorrelated between the neurons. Excitatory recurrent connections were mediated by slow synaptic conductances, like those used to model *N*-methyl-D-aspartate (NMDA) receptors, which ensures that states with low firing rates are stable (Wang 1999). Indeed, the simulations show that even states with firing rates as low as 10 Hz can engage in rivalry alternations (Fig. 8A). The mean dominance duration as a function of the conductance (inputs) reproduces Levelt's propositions (Fig. 8B; see following text) and the distribution of their durations follow a skewed Gaussian (Fig. 8C).

As found for our rate model, alternations in the spiking neuronal network are not mediated by adaptation; rather, they are noise driven. There are two sources of activity fluctuations in the spiking network model: noise in the synaptic input arriving from background sources (external to the network), and variability from nonsynchronous, individual, synaptic inputs generated within the network. In the network described earlier, transitions were driven by the latter source of noise. We verified this by increasing the size of the network while proportionally scaling down the unitary synaptic conductances, such that the mean total conductance to each neuron was kept constant, alhough the size of its input fluctuations was reduced (external noise is kept fixed here). As the network size increased, there was a point at which alternations ceased (at 10,000 neurons; not shown), revealing that the cause of the transitions was internally generated noise in the form of spiking neuronal variability. In addition to identifying the cause of alternations, this result also shows that adaptation alone cannot produce alternations in our model because the adaptation modulation was not affected by this manipulation. This last result also indicates that larger networks would require a mechanism that maintains the internal spiking variability to ensure that switching would be maintained. This could be obtained by small amounts of externally correlated noise (Moreno et al. 2002; Moreno-Bote and Parga 2006; Renart et al. 2007; Zohary et al. 1994). Indeed, simulations of a large-scale network (10,000 neurons per population) showed alternations when a small fraction of the external noise (~1%) was the same for all neurons in a population (not shown). Note that the firing rate models

with added noise presented in the previous sections are thought to represent such large networks with correlated fluctuations that cannot be averaged out. Because it is not feasible to perform long enough simulations of very large spiking networks to obtain reliable statistics, we simulated mainly small networks and therefore used completely uncorrelated external noise.

An interesting property of the model is presented in Fig. 9, which shows the mean firing rate as a function of the stimulus strength for the dominant and suppressed populations during rivalry, as well as the mean firing rate of a single population under nonrivaling conditions (e.g., when the competing monocular stimulus is turned off in binocular rivalry). During nonrivaling conditions, the mean firing rate of the stimulated population increases with input strength. In contrast, during rivalry the mean firing rate of the dominant population shows little dependence on stimulation strength. Furthermore, the activity during rivalry is lower than that during nonrivaling conditions throughout the range of stimulation strength. A similar reduction in activity was also observed in the rate-based and energy models (data not shown). Therefore a lower activity during rivalry is a robust feature in all versions of our model. This prediction can be tested experimentally using functional magnetic resonance imaging (fMRI) or electrophysiology (see DISCUSSION).

Dependence of mean dominance durations at large input strengths

Our energy-based, population rates, and spiking network models produce Levelt's Proposition II in a wide range of input strengths (Figs. 2B and 8B). This behavior is robust in our models as long as the input strength to the affected eye is smaller than or similar to the input strength to the other, unaffected eye (that whose input strength, e.g., image's contrast is kept fixed). However, consideration of the situation when the input strength of the affected eye is made much larger than that of the other eye suggests that, at some point, a different behavior than that stated in Levelt's Proposition II must emerge. This is readily apparent when one considers the limiting case, when the image contrast to the affected (say, the right) eye has been raised well above that of the other eye. Clearly, at this point perception will be dominated by the image given to the right eye, and presumably its mean dominance duration must be much higher than that of the other (left) eye. Therefore at a point around or soon after the contrast of the right eye is increased above the (fixed) contrast of the left eye, the mean duration of the right eye must start increasing, in violation of Levelt's Proposition II. This is precisely the behavior produced in our models, as shown in Fig. 10A. As the input strength to population B is made larger than that to population A, the mean dominance duration of B increases much more than the mean dominance duration of A is reduced. Figure 10B provides an intuitive explanation of this result in terms of the effect of input strength on the energy function. The energy function in Eq. 1 is plotted for several values of g_B, all of them larger than g_A. As g_B increases, state A_{on} starts to lose stability because its energy well becomes shallower. Crucially, at the same time the energy well of state B_{on} becomes deeper, thus increasing its mean dominance duration. Recently, Brascamp et al. (2006) showed that the behavior described earlier is indeed observed experimentally in binocular rivalry, i.e., that there is a significant violation of Levelt's Proposition II so that as the contrast of the affected eye is increased well above that of the unaffected eye, the mean dominance durations of the former rise rapidly. The same authors also showed that this behavior is found in purely oscillator models, although more analysis about its robustness is required.

Discussion

The mechanisms by which perceptual alternations occur during binocular rivalry are not well understood, nor is it known whether there are commonalities (e.g., similar architectures) with the mechanisms that cause perceptual switching in other bistable perceptual phenomena. The work presented here shows that attractor networks, as a class of models, provide a plausible

framework to describe the dynamics of perceptual bistability. Our approach differs from most existing models of bistability, which assume the alternations are driven by some form of slow adaptation acting on the dominant population (Kalarickal and Marshall 2000; Lago-Fernandez and Deco 2002; Laing and Chow 2002; Lehky 1988; Matsuoka 1984; Stollenwerk and Bode 2003; Wilson 2003). In those models, the adaptation precludes the persistence of the dominant state over time. The threshold for switching and the activity state slowly drift toward each other and autonomously coalesce, leading to a switch. The oscillation between the two competing populations is the only stable state in the system. In contrast, in our model the competing states remain stable fixed-points at all times, and it is noise (e.g., the spiking variability observed commonly in vivo) that causes alternations in dominance. Thus alternations cease if noise is removed (Figs. 1 and 6), although in its presence the interplay between adaptation and noise sets the timescale of alternations. Finally, the same sources of noise in our model also cause the variability in dominance durations observed experimentally; i.e., there is no need to invoke an ad hoc assumption about the presence of noise to explain this variability.

Our model goes beyond previous energy-based conceptualizations of perceptual bistability (e.g., Haken 1994; Kanai et al. 2005; Kim et al. 2006; Riani and Simonotto 1994), by presenting a neurally plausible attractor model that behaves consistently with experimental findings, most notably the increase in alternation rate with stimulation strength observed in binocular rivalry (Levelt 1968; Fig. 2B). This behavior would not arise automatically in an attractor-based model but rather depends on the network architecture. In particular, if the effect of increasing stimulus strength in an attractor model was to deepen the energy well of the corresponding attractor, this would have the opposite outcome of lengthening of the durations the network spends in that attractor. Furthermore, comprehensive analysis of oscillator models revealed that they, too, produce dominance durations that increase with stimulus strength in large parts of parameter space (Shpiro et al. 2006). The different behavior in our model (a shortening of the time spent in the competing attractors with increasing stimulus strength) arises from introducing in the energy function terms coupling the attractors' energy barriers with the input strength (Fig. 2A). In the network architecture of our model this was realized by feeding into each local inhibitory population a signal equal to the total external input [either directly (Fig. 3A) or by a global excitatory pool (Fig. 3B)], which is gated (multiplied) by the activity level of the corresponding excitatory population. Our network thus illustrates a class of models in which the wiring can be dynamically modified, as opposed to being hard-wired. As for the question what brain region may act as the excitatory pool in Fig. 3B (i.e., compute the total strength of all sensory inputs), note that this need not be a cortical region. Broad tuning like that expected from this hypothesized region is more characteristic of subcortical structures, which receive projections from a multitude of sensory cortical areas, and therefore could compute the global signal our model requires and send it back to the cortical local inhibitory subpopulations as schematized in Fig. 3B.

Noise versus adaptation as possible causes of perceptual alternations

The success of our model in reproducing salient dynamical behaviors of perceptual bistability suggests that noise may be the primary cause of perceptual alternations in bistability. This contrasts with the prevalent view that perceptual alternations are caused by some form of adaptation or fatigue (e.g., Kalarickal and Marshall 2000; Lago-Fernandez and Deco 2002; Laing and Chow 2002; Lehky 1968; Matsuoka 1984; Stollenwerk and Bode 2003; Wilson 2003). It is therefore important to note that, although it is known that there are multiple forms and mechanisms of adaptation in the brain, in the specific context of bistability a direct link has not been established to point to adaptation as the primary cause of alternations. An important observation in this context is that there is no evidence for dependence between the durations of successive dominance periods (e.g., a tendency for shorter periods to follow particularly long periods or vice versa; Fox and Herrmann 1967; Lehky 1995; Necker 1832;

Rubin and Hupé 2004), as may be expected if adaptation played a major role in causing alternations. Thus if adaptation plays any role in causing the alternations it would have to involve mechanisms with a very fast reset, so that all trace of it is essentially gone soon after the system has switched to the competing percept. Also at time-scales longer than individual dominance durations, data from long trials (5–10 min) of several binocular rivalry and plaid stimuli reveals that both the durations' means and their variances remain remarkably stable over time (Rubin and Hupé 2004), again showing no evidence for buildup of adaptation over time.

The distinction between oscillator models (where adaptation is the cause of alternations and noise is inessential) and noise-driven attractor models (where adaptation is inessential for alternations) is conceptually useful. However, given the ubiquity of adaptation mechanisms in the nervous system, in reality bistable networks most likely contain some form(s) of adaptation and this, in turn, may affect some aspects of the alternations. Indeed, we included adaptation in both our rate-based and spiking neuronal networks. Note, however, that in isolation (i.e., without noise) adaptation could not instigate alternations in our model and its function was rather to produce distributions of durations that resemble the skewed Gaussians observed experimentally. Specifically, the weak adaptation provided a natural and theoretically tractable way to implement a form of short-term persistence that disfavored the system leaving the attractor state right after it has settled into it, compared with the probability of leaving it later in time. However, other ways to implement such a tendency may be equally valid, such as synaptic facilitation of local inhibition by the selective excitatory population.

A few experimental studies examined the role of adaptation in perceptual bistability. Blake et al. (1990) modified the standard binocular rivalry paradigm by “forcing” one eye to dominate for long periods of time (30 s); they found that, on removal of the forcing, this eye's dominance durations were shorter by about a factor of two. Although this implicates adaptation in the dynamics, the crucial question is not whether adaptation is present, but whether it is responsible for the alternations. If that were the case, then the very long forcing should have led to very fast or even instantaneous transitions, with narrowly distributed durations. Instead, Blake et al. (1990) observed durations with a mean of about 2 s and large variability, suggesting that even such saturated adaptation does not instigate immediate transitions. In another experiment Leopold et al. (2002) showed, for a host of bistable stimuli, that alternations can be slowed down dramatically if the stimulus is periodically removed from view, again suggesting that if adaptation is involved in bistability, its influence does not carry over from one dominance epoch to another. Moreover, as these authors noted, their results suggest the involvement of a short-term implicit perceptual memory that, as discussed earlier, could produce distributions of dominance duration consistent with experiments without the need to invoke adaptation.

Two recent studies provide experimental evidence for an important role for noise in causing perceptual alternations. Brascamp et al. (2006) studied the role of noise in causing alternations by focusing on the prevalence of “return transitions,” cases when the dominant percept gives way to a mixed percept but then the system returns to the same percept that was dominant before (rather than the competing one). Brascamp et al. (2006) found a high prevalence of such transitions that, as they noted, is more consistent with noise than with adaptation as driving the alternations. Kim et al. (2006) studied the effect of weak contrast oscillations on the alternation rate of two rivaling images. They found the presence of stochastic resonance, that is, a maximum effect of the frequency of the oscillatory signal when it matches that of the alternations, an effect that can be explained only if a large amount of noise is present in the system.

The nature and sources of noise

In all three levels of description, the noise was fast compared with the timescale of alternations [*O*(100 ms) vs. *O*(1 s), respectively]; i.e., the transitions between states were not a trivial consequence of noise at the same scale. The noise timescale we posited is plausible biologically. In the spiking neuronal network, recurrent connections are dominated by NMDA-like synaptic receptors, and therefore the current fluctuations inherit the timescale of those synapses, of the order of 100 ms (Moreno-Bote and Parga 2005b; Titz and Keller 1997; Umemiya et al. 1999). Such receptors have been invoked in other models, e.g., to stabilize sustained activity in prefrontal cortex during a delayed-match-to-sample task (Wang 1999), and to account for the slow ramping behavior of neurons in posterior parietal cortex during a discrimination task (Wang 2002). Although in our simulations fast AMPA noise is present as an external source, internally generated noise is dominated by slower NMDA-generated fluctuations so as not to lead to fast population activity fluctuations that could destabilize the attractor dynamics.

There are other conceivable sources of noise in the synaptic input to cortical neurons. Modulations in ongoing cortical activity patterns, measured with optical imaging and local field potential recordings, are known to affect spiking responses in single neurons (Arieli et al. 1996). Moreover, although the underlying mechanisms are not well understood, changes in the level of coherent cortical activity at those timescales have been tied to modulations in attention and perceptual performance (Fries et al. 2001; Salinas and Sejnowski 2001; Womelsdorf et al. 2006). Thus variability appearing as mere noise in the context of perceptual bistability may arise from changes in internal network states that have functional roles in other situations. Finally, it is reasonable to assume that other sources of internal noise, including some that act at slower timescales (e.g., variations in global neurotransmitter levels, endogenous attention modulations, blinks) could also play a role in producing some of the switches.

Model predictions and experimental tests

An important feature of our model that has arisen from the energy formulation is the presence of inhibition from the input layer that is targeted at the competing population(s) (directly, as in Fig. 3A, or by an excitatory pool, as in Fig. 3B). This leads to a new prediction that can be tested with electrophysiological and neuroimaging studies. The model predicts that activity during rivalry should be lower compared with when the neural population receives the same input under nonrivaling conditions (Fig. 8). (Note that we use the term “rivalry” here in the general sense of two competing interpretations of a stimulus and, correspondingly, two rivaling neural populations. Therefore the prediction is not restricted to binocular rivalry but also holds for other bistable perceptual phenomena.) The reason is that during rivaling stimulation, local inhibition is enhanced due to the higher signal from the external input, leading to a reduction in the activity of the dominant excitatory population. Furthermore, the difference in activity between the two conditions grows as the stimulation strength increases (Fig. 9). In contrast, this prediction does not arise for models in which the dynamics is governed by adaptation currents (oscillator models). There, when a population becomes dominant, it receives no inhibitory inputs (because the only possible source is the suppressed population), and therefore no reduction of activity is expected compared with when the competing stimulus is turned off. The predictions that oscillator and our attractor model make in this regard are clearly different and could be used to determine which model better describes the neuronal dynamics during rivalry.

Interestingly, recent human fMRI studies of binocular rivalry provide some evidence to support the prediction of our attractor model. Reduced blood oxygenated level–dependent signal during rivalry compared with nonrivaling vision have been shown in the lateral geniculate nucleus (Haynes et al. 2005; Wunderlich et al. 2005) and visual cortical areas V1 through V4 (Lee and

Blake 2002; Polonsky et al. 2000). In higher visual areas, an fMRI study found no differences in activity between rivalry and nonrivalry conditions in the fusiform face and parahippocampal place areas (Tong et al. 1998), whereas electrophysiological recording in monkeys have shown reduced activity during rivalry in inferotemporal cortex and the superior temporal sulcus (Sheinberg and Logothetis 1997). Further investigation is therefore needed to examine this issue across the brain and for different bistability phenomena.

In conclusion, we have proposed a novel framework to model the bistable perceptual alternations observed during exposure to ambiguous or rivaling sensory stimuli. Our approach is based on the assumption that each of the competing percepts corresponds to a neuronal stable state, and the transitions between them are caused by noise. This differs from the prevalent view that the transitions are caused by an adaptation or fatigue process, which implies that the alternations reflect a limit cycle (oscillations) in neuronal state space. Starting from an energy-based model chosen to meet specific experimentally observed characteristics, we derived neurally plausible rate-based and spiking attractor neuronal networks, which are the first implementations of this broad class of models showing a dynamical behavior consistent with salient properties of perceptual bistable phenomena. Our results suggest that the hypothesis that competing percepts may correspond to the activation of different attractor states of neural activity, and that alternations between them may be driven by noise, is sustainable from a theoretical point of view, and should be examined experimentally with more care.

Acknowledgments

We thank S. Seung, A. Shpiro, and H. Sompolinsky for useful comments and discussions.

Grants: This work was supported by National Eye Institute Grant EY-14030 to N. Rubin and by a Swartz Foundation grant to N. Rubin and J. Rinzel.

Appendix A: Energy Model

This model is defined by a two-well energy function (Eq. 1). The variable Δr evolves according to Eq. 2 with time constant $\tau = 10$ ms. The noise $n(t)$ is an Ornstein–Uhlenbeck process (Risken 1989) with zero mean and deviation σ ($\sigma = 0.7$)

$$\frac{d}{dt}n = -\frac{n}{\tau_s} + \sigma\sqrt{\frac{2}{\tau_s}}\xi(t) \tag{A1}$$

where $\tau_s = 100$ ms and $\xi(t)$ is a white noise process with zero mean and $\langle \xi(t)\xi(t') \rangle = \delta(t - t')$. See numerical procedures in APPENDIX E.

Appendix B: Rate-Based Models

Model with direct cross-inhibition

Based on the single-variable energy function of Eq. 1, we formulate an energy function for a network with two populations, *A* and *B*, that are firing at rates r_A and r_B, respectively. This energy function

$$E(r_A,r_B)= -\tfrac{1}{2}(\alpha r_A+\alpha_B -2\beta r_A r_B) + \tfrac{1}{2}g_A\left[(1-r_A)^2+r_B\right]+\tfrac{1}{2}g_B\left[(1-r_B)^2+r_A\right] + \sum_{i=A,B}\int_0^{ri} f^{(-1)}(u)\,\mathrm{d}u \tag{B1}$$

has quadratic potentials placed at the states (r_A, r_B) = (1,0) and (0, 1). For simplicity, the firing rates are dimensionless here, measured in relation to a maximum firing rate so that $0 \le r_A, r_B \le 1$. Here, $f^{(-1)}$ denotes the inverse function of a neuronal population's input–output relation [i.e., firing rate = f(input)]. For weak stimuli and with f idealized as a step function [i.e., $f(u)$ = 0 for $u < 0$, and $f(u)$ = 1 elsewhere], the energy function is simply $E(r_A,r_B)= -(\alpha r_A^2+\alpha r_B^2 -2\beta r_A r_B)/2$;if $\beta>\alpha$, it has two minima, placed at (r_A, r_B) = (1, 0) and (0, 1), within the plane [0, 1] × [0, 1], where the firing rates are defined. Thus in this parameter range the inhibition each population exerts on the other is strong enough to preclude the two populations from being active at the same time.

The following network dynamics

$$\begin{aligned}\tau\frac{\mathrm{d}}{\mathrm{d}t}r_A &= -r_A+f\left[\alpha r_A-\beta r_B+g_A-(g_A+g_B)r_A\right]\\ \tau\frac{\mathrm{d}}{\mathrm{d}t}r_B &= -r_B+f\left[\alpha r_B-\beta r_A+g_B-(g_A+g_B)r_B\right]\end{aligned} \tag{B2}$$

minimizes the energy function $E(r_A, r_B)$. Strictly, $E(r_A, r_B)$ is a Lyapunov function for the dynamics defined in Eq. B2. That is, E is nonincreasing along trajectories

$$\frac{\mathrm{d}}{\mathrm{d}t}E(r_A,r_B)=\sum_{i=A,B}\frac{\mathrm{d}}{\mathrm{d}t}r_i\frac{\partial}{\partial r_i}E(r_A,r_B)\le 0 \tag{B3}$$

assuming that f is a nondecreasing function and $\tau > 0$ (Hertz et al. 1991).

Once stochastic input terms n_A and n_B are added, the dynamics produces transitions between the two local minima of the potential function, according to the equation set

$$\begin{aligned}\tau\frac{\mathrm{d}}{\mathrm{d}t}r_A &= -r_A+f\left[\alpha r_A-\beta r_B+g_A-(g_A+g_B)r_A+n_A\right]\\ \tau\frac{\mathrm{d}}{\mathrm{d}t}r_B &= -r_B+f\left[\alpha r_B-\beta r_A+g_B-(g_A+g_B)r_B+n_B\right]\end{aligned} \tag{B4}$$

The two noise terms are taken to be independent, continuous random processes as in the single-variable energy-based model (Eq. A1).

Model with inhibition driven indirectly by an excitatory pool and with weak adaptation

Here we derive the rate-based model for the architecture in Fig. 3B, with an excitatory pool projecting to all local inhibitory populations to mediate the mutual exclusivity, instead of direct cross-inhibition between the percept-specific populations. Despite the differences in architecture between this and the previous model, they can be mapped one onto the other for

particular parameters sets, as shown in the next section. Parameters below were chosen to allow this mapping.

The activity of the excitatory population A, r_A, is described by the equation set

$$\tau \frac{\mathrm{d}}{\mathrm{d}t} r_A = -r_A + f\left(\alpha r_A - \beta r_{A,\mathrm{inh}} + g_A - \alpha_A + n_A\right)$$
$$\tau_a \frac{\mathrm{d}}{\mathrm{d}t} \alpha_A = \alpha_A + \gamma r_A \tag{B5}$$

with $\alpha = 0.75$, $\beta = 0.5$, $\gamma = 0.1$; f is the input–output curve, modeled as a sigmoid function

$$f(x) = \langle 1 + \exp\{-(x-\theta)/k\}\rangle^{-1} \tag{B6}$$

with threshold $\theta = 0.1$ and $k = 0.05$. The inputs to the neuron consist of: recurrent excitation (with strength or efficacy α); local inhibition (with strength β) that grades linearly with the inhibitory firing rate $r_{A\ inh}$; a hyperpolarizing current, $a_A(t)$, with a maximum amplitude γ and time constant $\tau_a = 2$ s that produces weak adaptation; and the noise variable $n_A(t)$, with SD $\sigma = 0.03$.

The local inhibitory population A is assumed to respond instantaneously to its inputs (i.e., fast recruitment) with a quadratic input-output relation. The quadratic form allows the system to be easily mapped onto the previous architecture (see following text), although other steep nonlinear functions (e.g., cubic) also produce similar network behavior. Its firing rate is given by

$$r_{A,\mathrm{inh}} = (r_{\mathrm{pool}} + \eta r_A)^2 \tag{B7}$$

where $\eta = 0.5$ is the ratio between the strength of the excitatory feedback (see Fig. 3) and the input from the excitatory pool, r_{pool}.

The excitatory pool receives inputs from the network (weighted by $\varphi = 0.5$) and from the external stimulation, and we assume that it responds with a short recruitment timescale and linearly in response to its inputs. Its firing rate is therefore given by

$$r_{\mathrm{pool}} = [\varphi(r_A + r_B) + g_A + g_B]^+$$

where $[\bullet]^+$ denotes linear thresholding (note that the rate of the pool is nonnegative even when inputs are negative, allowing one to define the system also in that input regime). Similar equations define the dynamics of the population selective to stimulus B.

Relation between the models with and without direct cross-inhibition

Despite the large differences in the architecture between the two rate-based models we have presented, it is possible to map approximately one into the other for particular sets of parameters. In fact, parameters in the model without cross-inhibition have been chosen to allow this mapping, and therefore to have dynamics consistent with Levelt's propositions. We next explain the mapping.

Let us start from the case with no direct inhibition. Because the response properties of interest are found with small stimulation strengths ($g_{A,B} \ll 1$), we may approximate the activity of the local inhibitory population (Eq. B7) by

$$r_{A,\mathrm{inh}} \approx [(\varphi+\eta)r_A+\varphi r_B]^2+2(g_A+g_B)[(\varphi+\eta)r_A+\varphi r_B]$$

During rivalry alternations values of $r_{A,B}$ are either close to 0 or 1 because the input–output relation for the excitatory population is rather steep and saturates (Eq. B6). Suppose that population A is dominant, so that B is inactive ($r_B \sim 0$); then

$$r_{A,\mathrm{inh}} \approx (\varphi+\eta)^2 r_A^2+2(\varphi+\eta)(g_A+g_B)r_A$$

Therefore the dynamics of r_A when B is suppressed are approximately governed by

$$\tau\frac{\mathrm{d}}{\mathrm{d}t}r_A = -r_A+f[\alpha r_A-\beta(\varphi+\eta)^2 r_A^2+g_A -2\beta(\varphi+\eta)(g_A+g_B)r_A-\alpha_A+n_A] \quad \text{(B8)}$$

Now compare this equation with the corresponding Eq. B4 from the case with direct inhibition. If we set $2\beta(\varphi+\eta)=1$ (or approximately so) both equations depend identically on the stimulation strengths. Although the term $\beta(\varphi+\eta)^2 r_A^2$ differs between the two models, this difference does not affect the qualitative behavior as the stimulus strengths are varied. However, this term imposes a number of conditions that should hold to allow alternations and to produce mutual exclusivity between the possible stationary states. Because the dominance state should be stable, we have to impose the condition that the total synaptic input to the population is on average above the firing threshold, that is $\alpha-\beta(\varphi+\eta)^2>\theta$. To guarantee mutual exclusivity we demand that if both populations attempt to become active simultaneously, the net input should be below threshold. This condition imposes $\alpha-\beta(2\varphi+\eta)^2<\theta$. The preceding three conditions should hold to produce a dynamic properties that are consistent with experimental observations. In simulations we have chosen $\alpha=0.75$ and $\beta=\varphi=\eta=0.5$, although others are also valid.

Besides the architecture, the main difference between the two models is the presence of adaptation for the model with indirect inhibition. Adaptation shapes the distribution of dominance durations but its influence is limited; we choose parameter values such that adaptation by itself (without noise) does not generate transitions between percepts. This means that the activity of a fully adapted dominant population cannot drop below the threshold of the input-output relation. Because the maximum amplitude of adaptation is γ (see Eq. B5), this condition translates into the parameter constraint (from Eq. B8) $\alpha-\beta(\varphi+\eta)^2-\gamma>\theta$, We have used $\gamma=0.1$ and, to generate alternations with a duration of a few seconds, we have chosen $\sigma=0.03$, unless noted otherwise.

Appendix C: Spiking Neuronal Network

We have simulated a cell-based network with the connectivity described in Fig. 3B. Each population contains $N=100$ leaky integrate-and-fire neuron models. Coupling is with conductance-based synapses and all-to-all connectivity (each neuron receives connections from *all* neurons in a presynaptic population). Model equations and parameters follow (Brunel

and Wang 2001;Moreno-Bote and Parga 2005a,b;Wang 2002). The voltage below the spiking threshold for the excitatory neurons in the competing populations obeys

$$C_m \frac{\mathrm{d}}{\mathrm{d}t} V(t) = -g_L[V(t) - V_L] - I_{\mathrm{syn}}(t) - I_{\mathrm{adop}}(t)$$

with membrane capacitance $C_m = 0.5$ nF, leak conductance $g_L = 25$ nS, producing a membrane time constant $\tau_m = C_m/g_L = 20$ ms, and resting potential $V_L = -65$ mV. The neuron emits a spike when the voltage reaches the threshold $V_{th} = -54$ mV, after which the voltage is reset to $V_{reset} = -60$ mV. $I_{syn}(t)$ is the total synaptic current delivered to a neuron. $I_{adap}(t)$ is a slow conductance-driven adaptation current: $I_{adap}(t) = g_{adap}(t)(V - V_{adap})$; g_{adap} is increased by $\Delta g = 0.075$ nS with each spike and decays to zero exponentially with time constant $\tau_{adap} = 2$ s; $V_{adap} = -80$ mV. Voltage equations for the inhibitory populations and the pool are the same, but without adaptation current.

The synaptic currents to the excitatory (E), inhibitory (I) populations, and the pool (P) are $I_{syn,E}(t) = I_{NMDA,rec}(t) + I_{GABA}(t) + I_{ext,E}(t) + I_{back}(t)$, $I_{syn,I}(t) = I_{AMPA}(t) + I_{ext,I}(t) + I_{back}(t)$, and $I_{syn,P}(t) = I_{AMPA}(t) + I_{ext,P}(t) + I_{back}(t)$, respectively. The NMDA recurrent synaptic current is modeled with a linearized driving force (valid below threshold) as $I_{\mathrm{NMDA,rec}} = \sum_i^N g_{\mathrm{NMDA},i}(t)(V - V_E)$ (Brunel and Wang 2001; Moreno-Bote and Parga 2005b), where $V_E = 0$ mV and $g_{NMDA,i}(t)$ is the individual conductance generated by the presynaptic neuron i, defined as

$$\frac{\mathrm{d}}{\mathrm{d}t} g_{\mathrm{NMDA},i}(t) = -\frac{g_{\mathrm{NMDA},i}(t)}{\tau_{\mathrm{NMDA}}} + [g_{\mathrm{NMDA,max}} - g_{\mathrm{NMDA},i}(t)] \sum_j \delta(t - \mathrm{t}_j^i)$$

Here, $\tau_{NMDA} = 100$ ms, $g_{NMDA,\mathrm{max}} = 0.15$ nS is the individual synaptic maximum conductance, and the sum represents the spikes emitted by neuron i at previous times t_j^i. Equations for the AMPA and GABA currents are (k = AMPA, GABA) given by $I_k = g_k(V - V_k)$, where the conductance is

$$\frac{\mathrm{d}}{\mathrm{d}t} g_k(t) = -\frac{g_k(t)}{\tau_k} + g_{\mathrm{pop,unit}} \sum_{j,i} \delta(t - \mathrm{t}_j^i)$$

with $\tau_{AMPA(\mathrm{GABA})} = 10$ ms (20 ms), $V_{E(\mathrm{I})} = 0$ mV (-80 mV), and the sum of spikes now extends to all presynaptic neurons i. The unitary conductance for E to P connections is $g_{E\to\mathrm{P,unit}} = 0.075$ nS, $g_{P\to\mathrm{I,unit}} = 0.23$ nS for the P to inhibitory population (I) connections, $g_{I\to\mathrm{E,unit}} = 0.175$ nS for the I to E connections, and $g_{E\to\mathrm{I,unit}} = 0.1$ nS for the E to I synapses.

External inputs are modeled as constant excitatory conductances to produce a current $I_{ext,E,\mathrm{A(B)}} = g_{A(\mathrm{B})}(V - V_E)$ for the E populations A(B), $I_{ext,I,\mathrm{A(B)}} = f_I g_{A(\mathrm{B})}(V - V_E)$ for the I populations A(B), and $I_{ext,P} = f_P(g_A + g_B)(V - V_E)$ for P, following the architecture of Fig. 3B. The factors $f_I = f_P = 0.1$ measure the effect of the external inputs on I and P populations in relation to E, and they control the slope in Fig. 9.

Each neuron receives an independent source of noisy conductance with AMPA and GABA contributions mimicking spontaneous external activity (Destexhe et al. 2003; Moreno-Bote and Parga 2005a), defined by $I_{\text{back}}(t)=\sum_{k\,(k=\text{AMPA,NMDA})} [\,g_k+n_k(t)][\,V(t)-V_k\,]$, where $n_k(t)$ is a colored noise as in *Eq. 3* with timescale $\tau_{AMPA(\text{GABA})}$ = 10 ms (20 ms). The means (g_k) and dispersions (σ_k) for the background conductances are $g_{AMPA(\text{GABA})}$ = 5 nS (7.5 nS), $\sigma_{AMPA(\text{GABA})}$ = 3.53 nS (3.53 nS), equal for all neurons.

The parameters are chosen as follows: Background conductances alone should produce low firing rates in all populations. Connections between E and P, and P to I should be strong enough to produce winner-take-all behavior. Recurrent NMDA connections should be tuned to support attractor states and also allow transitions between them. Connections between E and I also need to be strong.

Appendix D: Log-Normal and Gamma Fits to the Distribution of Dominance Durations

The distributions of dominance durations in Fig. 7A have been fitted with log-normal and gamma distributions. The log-normal distribution is defined as

$$f_{\log n}(t)=\frac{1}{\sqrt{2\pi\sigma}}\exp\left\{-(\log t-\mu)^2/2\sigma^2\right\}$$

and the gamma distribution as

$$f_{\text{gamma}}(t)=Ct^{\alpha-1}e^{-t/\beta}$$

where C is the normalizing factor. Maximum likelihood fits of the simulated distributions with a log-normal distribution give the values μ = 1.24 and σ = 0.35, and with the gamma distribution give α = 8.66 and β = 0.41. The quality of both fits is very similar.

Appendix E: Numerical Procedures

The dynamical equations for energy, rate-based, and spiking network simulations are integrated using Euler's method with time step δt = 0.1 ms. Recomputing with a shorter integration time step did not produce appreciable differences in any of the results that we obtained and reported with the standard time step. The dominance durations for each percept in the energy model are defined by the amount of time in which the variable Δr is below (or above) Δr = 0. For the rate-based model, a transition occurs when the firing rate becomes larger (or smaller) than the firing rate of the other population. For the spiking network, a transition occurs when the averaged population firing rate (number of spikes emitted by the excitatory population over time window Δt = 100 ms divided by the number of excitatory neurons) reverses order with the firing rate of the competing population. In this case, due to the large activity fluctuations, we impose the additional constraint in defining a transition that the firing rate of the population that becomes dominant must be at least >5 Hz. Energy and rate-based models typically run for 10^5 s (model time), generating around 10^4 durations for each percept. Means in all the plots are computed from the time series generated with these long simulations and error bars correspond to SDs of the means. Also the distributions and switch-triggered averages obtained from these time series are smooth and robust. For the spiking network simulations, shorter runs

were used due to the large number of neurons per population being simulated. These spiking network simulations typically run for 10^4 s, producing on the order of 10^3 alternations. We used Fortran 90 custom code to simulate the models, and Matlab to analyze and plot the data, along with a random generator for white noise that generated long nonrepetitive series.

References

Arieli A, Sterkin A, Grinvald A, Aertsen A. Dynamics of ongoing activity: explanation of the large variability in evoked cortical responses. Science 1996;273:1868–1871. [PubMed: 8791593]

Bialek W, DeWeese M. Random switching and optimal processing in the perception of ambiguous signals. Phys Rev Lett 1995;74:3077–3080. [PubMed: 10058097]

Blake R. A neural theory of binocular rivalry. Psychol Rev 1989;96:145–167. [PubMed: 2648445]

Blake R. A primer on binocular rivalry. Brain Mind 2001;2:5–38.

Blake R, Westendorf D, Fox R. Temporal perturbations of binocular rivalry. Percept Psychophys 1990;48:593–602. [PubMed: 2270191]

Brascamp JW, van Ee R, Noest AJ, Jacobs RHAH, van den Berg AV. The time course of binocular rivalry reveals a fundamental role of noise. J Vision 2006;6:1244–1256.

Brunel N, Wang XJ. Effects of neuromodulation in a cortical network model of object working memory dominated by recurrent inhibition. J Comput Neurosci 2001;11:63–85. [PubMed: 11524578]

Dayan P. A hierarchical model of binocular rivalry. Neural Comput 1998;10:1119–1135. [PubMed: 9654769]

Destexhe A, Rudolph M, Paré D. The high-conductance state of neocortical neurons in vivo. Nat Rev Neurosci 2003;4:739–751. [PubMed: 12951566]

Fox R, Herrmann J. Stochastic properties of binocular rivalry alternations. Percept Psychophys 1967;2:432–436.

Freeman AW. Multistage model for binocular rivalry. J Neurophysiol 2005;94:4412–4420. [PubMed: 16148271]

Fries P, Reynolds JH, Rorie AE, Desimone R. Modulation of oscillatory neuronal synchronization by selective visual attention. Science 2001;291:1560–1563. [PubMed: 11222864]

Haken H. A brain model for vision in terms of synergetics. J Theor Biol 1994;171:75–85.

Haynes JD, Deichmann R, Rees G. Eye-specific effects of binocular rivalry in the human lateral geniculate nucleus. Nature 2005;438:496–499. [PubMed: 16244649]

Hertz, J.; Krogh, A.; Palmer, RG. Introduction to the Theory of Neural Computation. Redwood City, CA: Addison–Wesley; 1991.

Hupé JM, Rubin N. The dynamics of bi-stable alternation in ambiguous motion displays: a fresh look at plaids. Vision Res 2003;43:531–548. [PubMed: 12594999]

Kalarickal GJ, Marshall JA. Neural model of temporal and stochastic properties of binocular rivalry. Neurocomputing 2000:32–33. 843–853.

Kanai R, Moradi F, Shimojo S, Verstraten FAJ. Perceptual alternation induced by visual transients. Perception 2005;34:803–822. [PubMed: 16124267]

Kim YJ, Grabowecky M, Suzuki S. Stochastic resonance in binocular rivalry. Vision Res 2006;46:392–406. [PubMed: 16183099]

Kramers HA. Brownian motion in a field of force and the diffusion model of chemical reactions. Physica 1940;7:284–304.

Lago-Fernandez LF, Deco G. A model of binocular rivalry based on competition in IT. Neurocomputing 2002;44:503–507.

Laing CR, Chow CC. A spiking neuron model for binocular rivalry. J Comput Neurosci 2002;12:39–53. [PubMed: 11932559]

Lankheet MJ. Unraveling adaptation and mutual inhibition in perceptual rivalry. J Vis 2006;6:304–310. [PubMed: 16889470]

Lee SH, Blake R. V1 activity is reduced during binocular rivalry. J Vis 2002;2:618–626. [PubMed: 12678633]

Lehky SR. An astable multivibrator model of binocular rivalry. Perception 1988;17:215–228. [PubMed: 3067209]

Lehky SR. Binocular rivalry is not chaotic. Proc R Soc Lond B Biol Sci 1995;259:71–76.

Leopold DA, Logothetis NK. Multistable phenomena: changing views in perception. Trends Cogn Sci 1999;3:254–264. [PubMed: 10377540]

Leopold DA, Wilke M, Maier A, Logothetis NK. Stable perception of visually ambiguous patterns. Nat Neurosci 2002;5:605–609. [PubMed: 11992115]

Levelt, WJM. On Binocular Rivalry. Paris: Mouton; 1968.

Logothetis NK. A primer on binocular rivalry, including current controversies. Philos Trans R Soc Lond B Biol Sci 1998;353:1801–1818. [PubMed: 9854253]

Lumer ED. A neural model of binocular integration and rivalry based on the coordination of action-potential timing in primary visual cortex. Cereb Cortex 1998;8:553–561. [PubMed: 9758218]

Matsuoka K. The dynamic model of binocular rivalry. Biol Cybern 1984;49:201–208. [PubMed: 6704442]

Moreno R, de la Rocha J, Renart A, Parga N. Response of spiking neurons to correlated inputs. Phys Rev Lett 2002;89:288101. [PubMed: 12513181]

Moreno-Bote R, Parga N. Membrane potential and response properties of populations of cortical neurons in the high conductance state. Phys Rev Lett 2005a;94:088103. [PubMed: 15783940]

Moreno-Bote R, Parga N. Simple model neurons with AMPA and NMDA filters: role of synaptic time scales. Neurocomputing 2005b;65:441–448.

Moreno-Bote R, Parga N. Auto- and crosscorrelograms for the spike response of leaky integrate-and-fire neurons with slow synapses. Phys Rev Lett 2006;96:028101. [PubMed: 16486646]

Necker LA. Observations on some remarkable phenomenon which occurs on viewing a figure of a crystal of geometrical solid. Lond Edinburgh Philos Mag J Sci 1832;3:329–337.

Polonsky A, Blake R, Braun J, Heeger DJ. Neuronal activity in human primary visual cortex correlates with perception during binocular rivalry. Nat Neurosci 2000;3:1153–1159. [PubMed: 11036274]

Pressnitzer D, Hupé JM. Temporal dynamics of auditory and visual bistability reveal common principles of perceptual organization. Curr Biol 2006;16:1351–1357. [PubMed: 16824924]

Renart A, Moreno-Bote R, Wang XJ, Parga N. Mean-driven and fluctuation-driven persistent activity in recurrent networks. Neural Comput 2007;19:1–46. [PubMed: 17134316]

Riani M, Simonotto E. Stochastic resonance in the perceptual interpretation of ambiguous figures: a neural network model. Phys Rev Lett 1994;72:3120–3123. [PubMed: 10056072]

Risken, H. The Fokker-Planck Equation. Berlin: Springer-Verlag; 1989.

Rubin, E. Visuell wahrgenommene Figuren. Copenhagen: Gyldendals; 1921.

Rubin N. Binocular rivalry and perceptual multi-stability. Trends Neurosci 2003;26:289–291. [PubMed: 12798596]

Rubin, N.; Hupé, JM. Dynamics of perceptual bistability: plaids and binocular rivalry compared. In: Alais, D.; Blake, R., editors. Binocular Rivalry. Cambridge, MA: MIT Press; 2004.

Salinas E. Background synaptic activity as a switch between dynamical states in a network. Neural Comput 2003;15:1439–1475. [PubMed: 12816561]

Salinas E, Sejnowski TJ. Correlated neuronal activity and the flow of neural information. Nat Rev Neurosci 2001;2:539–550. [PubMed: 11483997]

Sheinberg DL, Logothetis NK. The role of temporal cortical areas in perceptual organization. Proc Natl Acad Sci USA 1997;94:3408–3413. [PubMed: 9096407]

Shpiro A, Rodica C, Rinzel J, Rubin N. Dynamical characteristics common to neuronal competition models. J Neurophysiol 2007;97:462–473. [PubMed: 17065254]

Stollenwerk L, Bode M. Lateral neural model of binocular rivalry. Neural Comput 2003;15:2863–2882. [PubMed: 14629871]

Suzuki S, Grabowecky M. Evidence for perceptual “trapping” and adaptation in multistable binocular rivalry. Neuron 2002;36:143–157. [PubMed: 12367513]

Titz S, Keller BU. Rapidly deactivating AMPA receptors determine excitatory synaptic transmission to interneurons in the nucleus tractus solitarius from rat. J Neurophysiol 1997;78:82–91. [PubMed: 9242263]

Tong F. Competing theories of binocular rivalry. Brain Mind 2001;2:55–83.

Tong F, Nakayama K, Vaughan JT, Kanwisher N. Binocular rivalry and visual awareness in human extrastriate cortex. Neuron 1998;21:753–759. [PubMed: 9808462]

Umemiya M, Senda M, Murphy TH. Behaviour of NMDA and AMPA receptor-mediated miniature EPSCs at rat cortical neuron synapses identified by calcium imaging. J Physiol 1999;521:113–122. [PubMed: 10562338]

van Ee R. Dynamics of perceptual bi-stability for stereoscopic slant rivalry and a comparison with grating, house-face, and Necker cube rivalry. Vision Res 2005;45:29–40. [PubMed: 15571736]

van Kampen, BF. Stochastic Processes in Physics and Chemistry. Amsterdam: North-Holland; 2001.

Wang XJ. Synaptic basis of cortical persistent activity: the importance of NMDA receptors to working memory. J Neurosci 1999;19:9587–9603. [PubMed: 10531461]

Wang XJ. Probabilistic decision making by slow reverberation in cortical circuits. Neuron 2002;36:955–968. [PubMed: 12467598]

Wheatstone C. Contributions to the physiology of vision. Part I. On some remarkable, and hitherto unobserved, phenomena of binocular vision. Lond Edinburgh Dublin Philos Mag J Sci 1838;3:241–267.

Wilson HR. Computational evidence for a rivalry hierarchy in vision. Proc Natl Acad Sci USA 2003;100:14499–14503. [PubMed: 14612564]

Womelsdorf T, Fries P, Mitra PP, Desimone R. Gamma-band synchronization in visual cortex predicts speed of change detection. Nature 2006;439:733–736. [PubMed: 16372022]

Wunderlich K, Schneider KA, Kastner S. Neural correlates of binocular rivalry in the human lateral geniculate nucleus. Nat Neurosci 2005;8:1595–1602. [PubMed: 16234812]

Zohary E, Shadlen MN, Newsome WT. Correlated neuronal discharge rate and its implications for psychophysical performance. Nature 1994;370:140–143. [PubMed: 8022482]

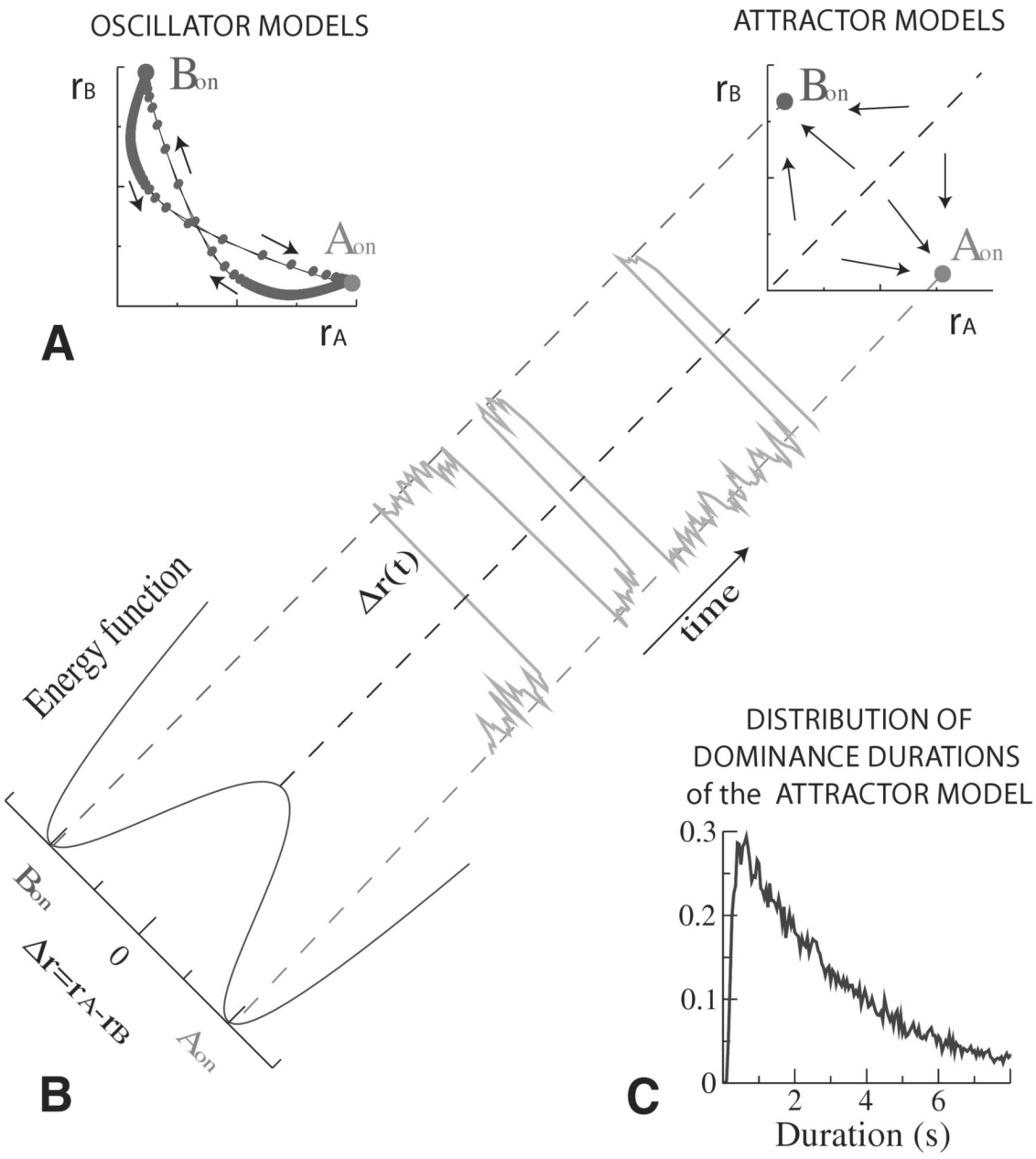

FIG. 1.

Oscillator vs. attractor models for bistability. *A, left*: in oscillator models, slow negative feedback (spike frequency adaptation or synaptic depression) produces periodic alternations between 2 states, A_{on} and B_{on}, seen as a closed trajectory on the plane of the population rates (r_A, r_B). *Right*: in attractor models, 2 possible steady states coexist, A_{on} and B_{on}. When the system is initialized in one or the other side of the separatrix (diagonal), it evolves and settles down in the closest attractor state. *B*: diagonally oriented panel shows the behavior of attractor models. Without noise, the system does not alternate; rather, trajectories lie on either side of the separatrix (dashed line) and approach one of the 2 states (attractors). In the presence of noise, alternations are produced. Time course of the difference between the 2 population firing rates, $\Delta r(t) = r_A - r_B$, is represented by the projection of a true random trajectory in the (r_A, r_B) plane onto a diagonal joining the points A_{on} and B_{on} (*middle*). *Bottom left*: energy function

underlying the dynamics of $\Delta r(t)$, which consists of 2 minima at A_{on} and B_{on}, and a local maximum at zero. *C*: distribution of dominance durations for the attractor model with equal input strengths.

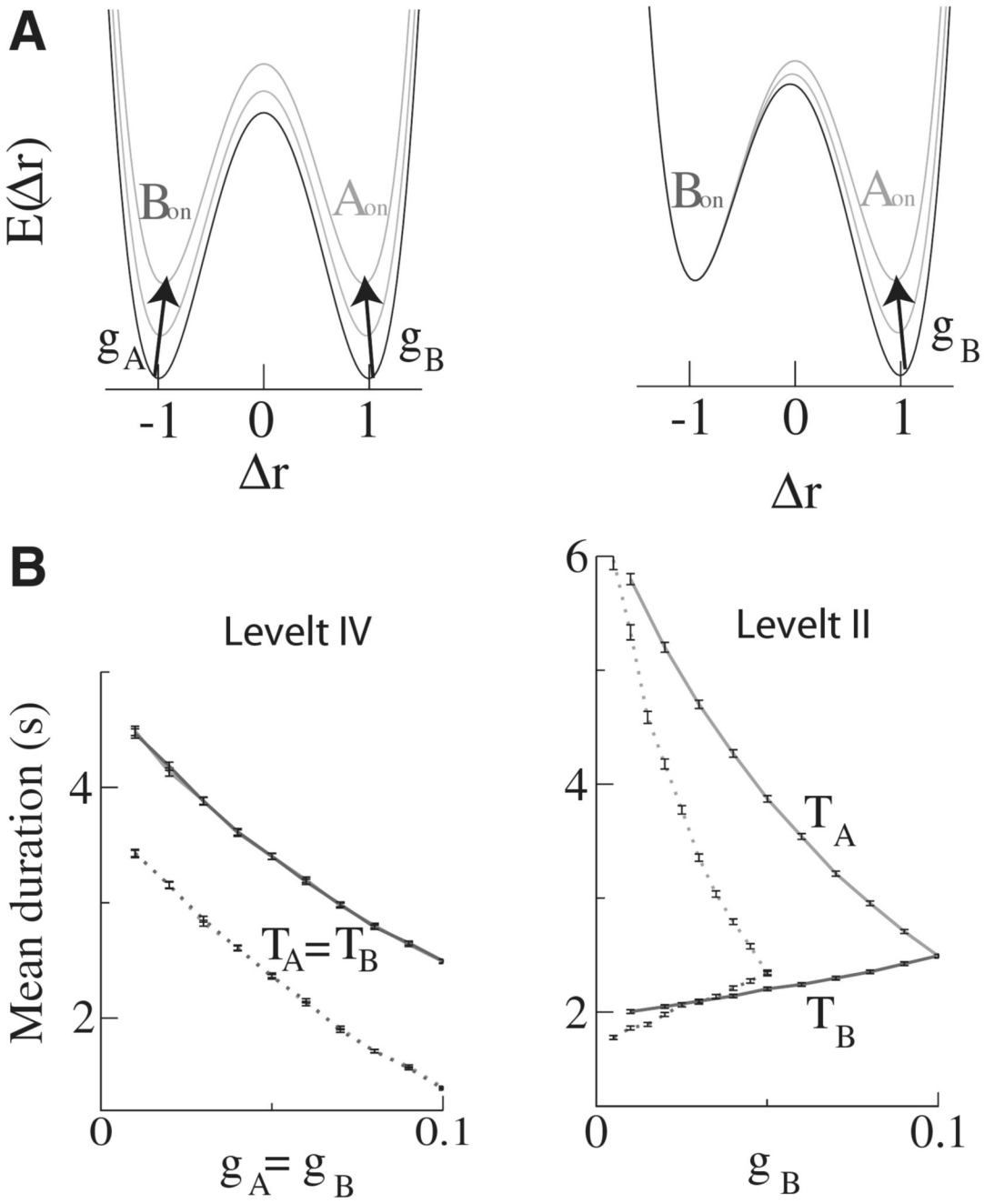

FIG. 2.

The bistable energy model produces Levelt propositions II and IV. *A*: energy functions when both stimuli are strengthened simultaneously, $g_A = g_B$ (*left*), and when g_A is kept fixed and g_B is varied (*right*). Energy wells change with g_B as indicated by the arrows. *B*: mean dominance durations for states A_{on} and B_{on}, denoted T_A and T_B, respectively, when the stimulation strengths vary simultaneously (*left;* Levelt's proposition IV) and when only g_B is varied (*right*; Levelt's proposition II). Full lines are for the one variable model in Eq. 2. Dashed lines are for the 2-variable rate-based model. *Right:* g_A is kept fixed at 0.1 and 0.05 for the one- and 2-variable models, respectively.

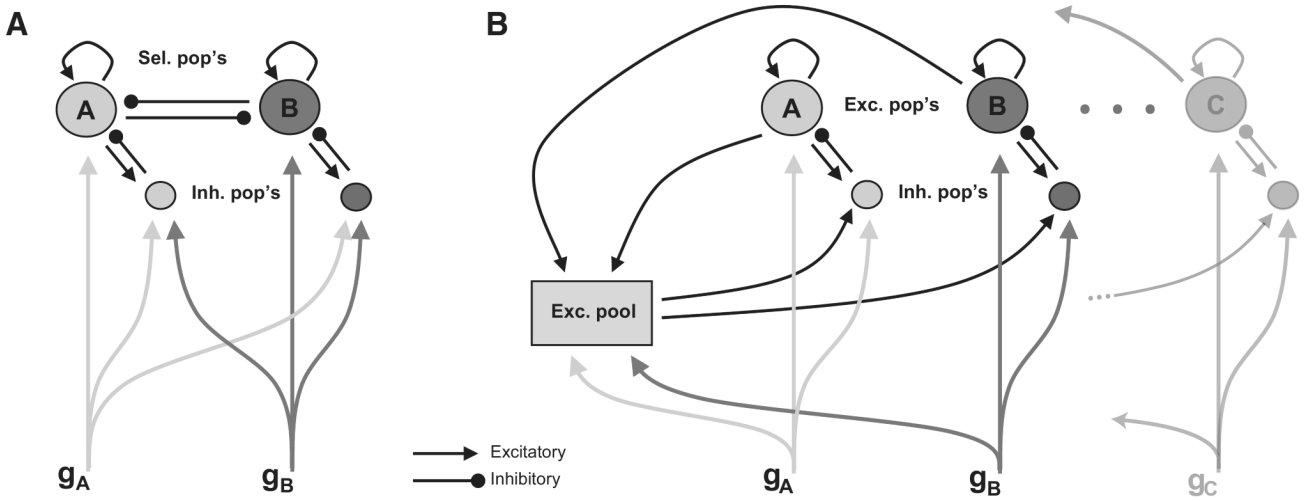

FIG. 3.

Architectures of network models for bistability. *A*: 2 recurrent neuronal populations that represent percepts *A* and *B* mutually inhibit each other directly. A prominent feature of this architecture is that separate local inhibitory subpopulations relay information about the total strength of external stimulation, $g_A + g_B$. *B*: for the architecture to generalize to multistability (between more than 2 competing percepts; here A, B, and C are shown), an excitatory pool is included that provides information about the total external stimulation to all local inhibitory subpopulations. This network does not require direct mutual inhibition between the competing populations because inhibition is delivered to them indirectly by feedback through the global excitatory pool.

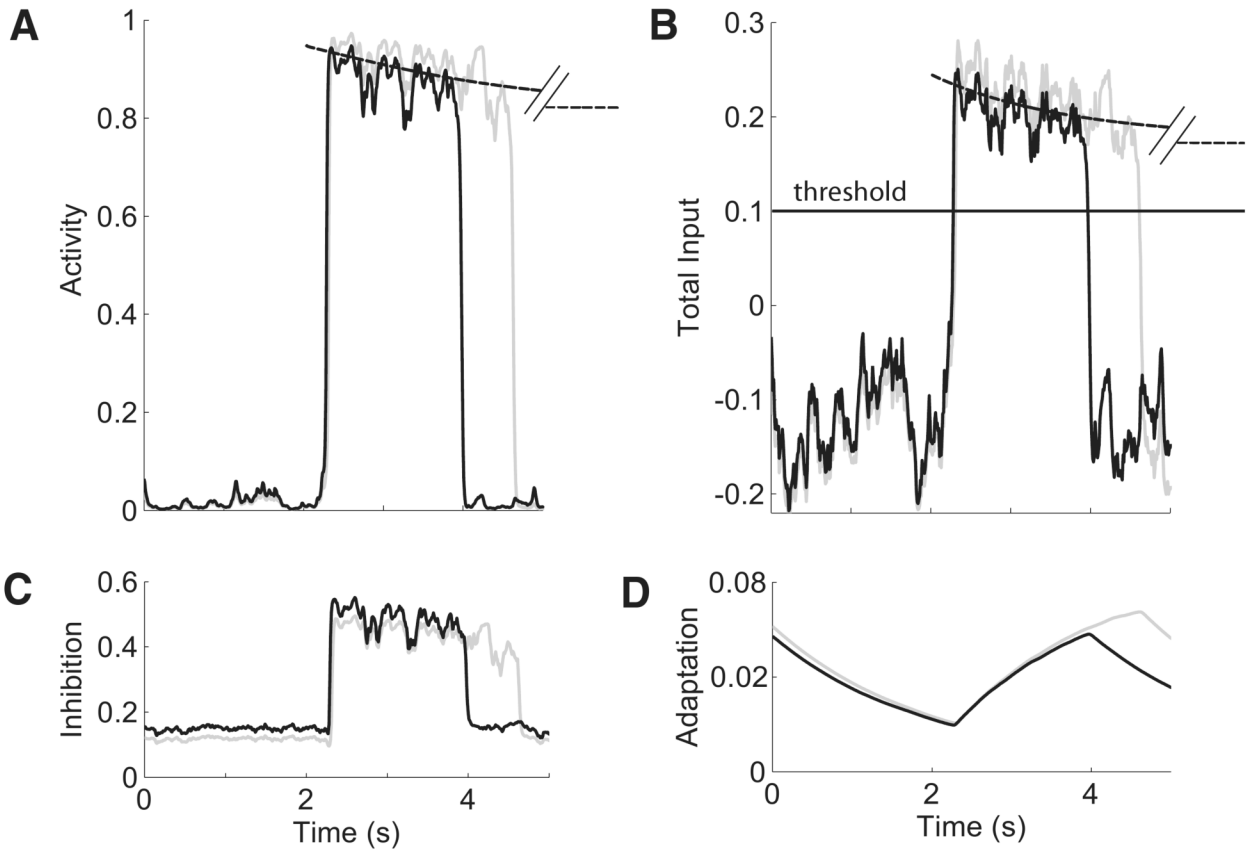

FIG. 4.

Dynamical properties of the 2-variable rate-based model. Time courses of dynamical variables of an excitatory neural population in Eqs. B5-B7. The population undergoes transitions between suppressed and dominant states. During dominance, its activity (*A*) and total input (*B*) are lower for strong stimulation ($g_{A,B}$ = 0.05, black curves) than for weak stimulation ($g_{A,B}$ = 0.01, gray curves). This reduction occurs due to stronger inhibition (*C*). Adaptation current (*D*) is affected by stimulus strength the same way as the activity. Horizontal line in *B* corresponds to the threshold (θ = 0.1) of the input–output transfer function of the excitatory neurons (Eq. B5). For the low-stimulation case, the dashed curves are the rate (*A*) and the total input (*B*); they correspond to the attractor's asymptotic or steady-state values for activity (*A*) and total current (*B*) in the absence of noise and given the instantaneous level of adaptation. Notice that this asymptotic level is well above the transition threshold (horizontal line) of the system. Identical external noise was used for the simulation in the 2 cases.

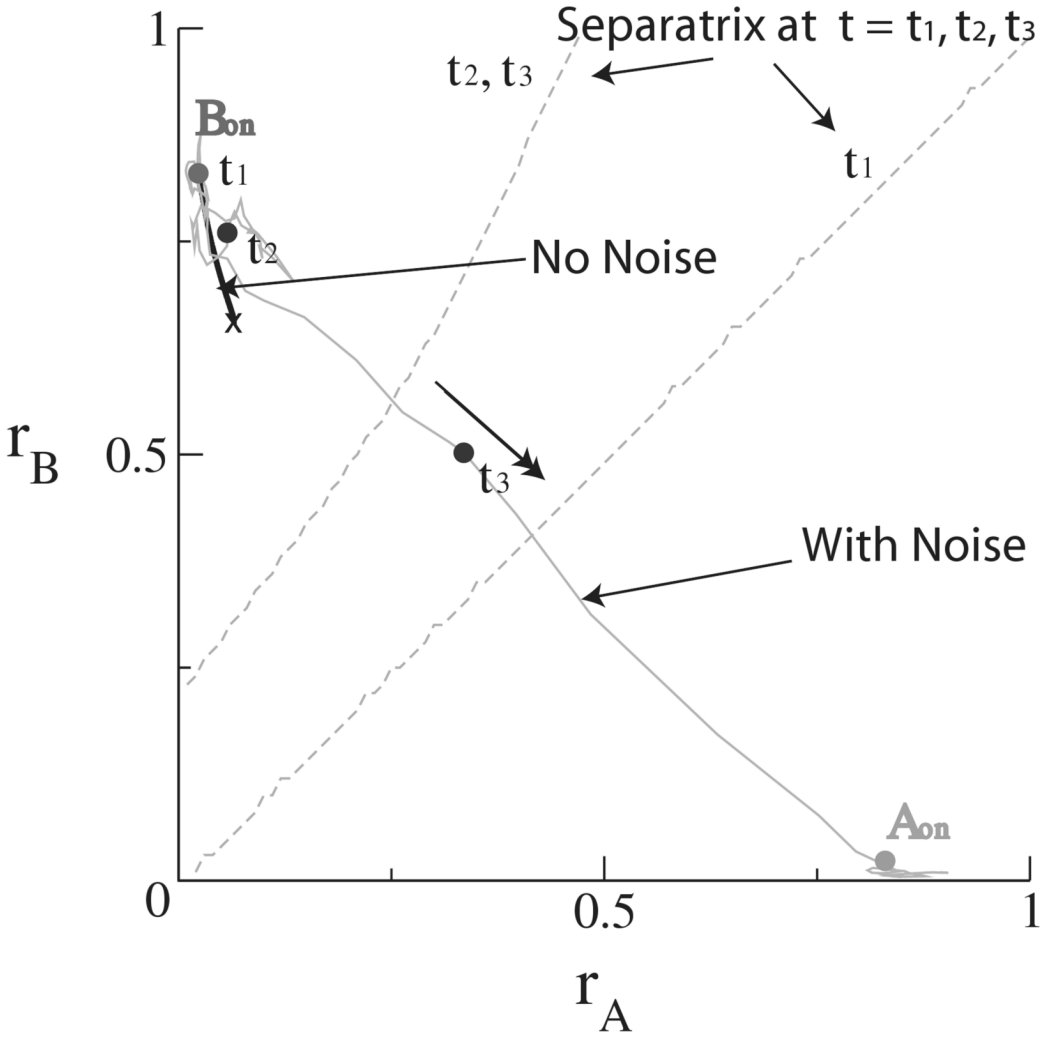

FIG. 5.

Adaptation alone, in a noise-free network, does not produce alternations. Smooth trajectory in the (r_A, r_B) phase plane starts from the initial point B_{on} and evolves to a steady-state point, marked by "×"; the trajectory does not reach the separatrix (diagonal line t_1) and therefore no alternation occurs. When noise is added, the trajectory (wiggly curve) wanders around B_{on} passing through the points t_1, t_2; at the same time the separatrix is drifting leftward (due to adaptation). Strong adaptation would have led to a transition even in the absence of noise. However, in our model the adaptation is weak and the separatrix remains far from B_{on}. Therefore a large noise fluctuation is needed to cross it; once this occurs, around point t_3, a switching to state A_{on} occurs; the switch occurs very rapidly and therefore the trajectory looks smooth during the transition. This geometrical representation illustrates that, to induce a switch to dominance in *A*, perturbations that are orthogonal to the separatrix are more effective, i.e., perturbations that increase r_A and reduce r_B simultaneously (see also Fig. 6B). Value of the stimulation was equal for both populations (g = 0.1).

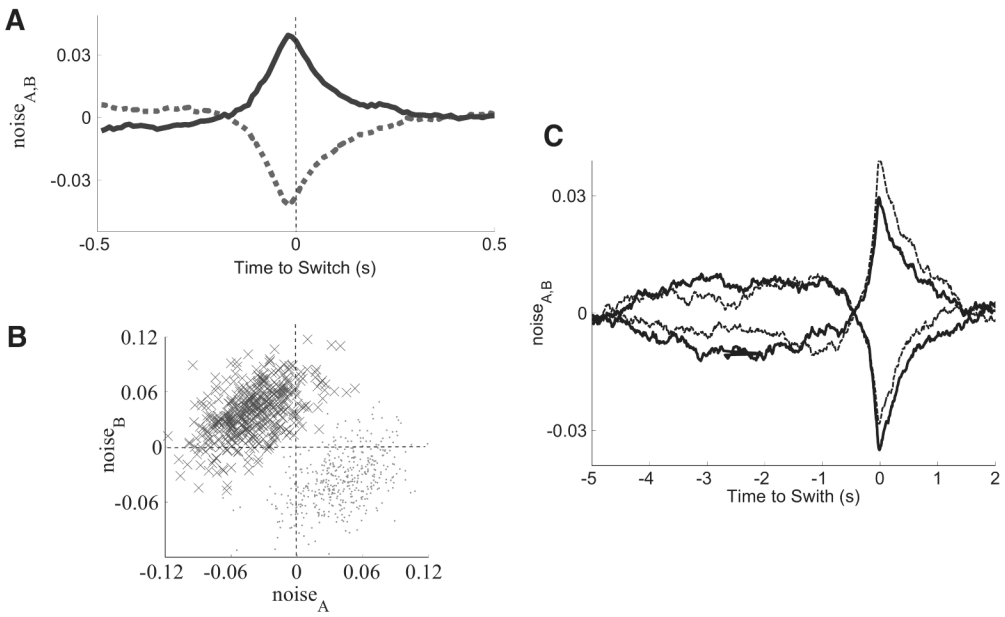

FIG. 6.

Noise directly drives alternations in the rate-based model. *A*: switch-triggered averages (STAs) of time courses of the input noise $n(t)$, time locked to network transition events. Solid curve is for transitions from the suppressed to the dominant state; the dashed curve is for the reverse. Switches are associated with simultaneous occurrence of lower-than-average noise to the dominant population and higher-than-average noise to the suppressed population. *B*: values of the noise input at moments of a transition ($t = 0$ in *A*) when population A becomes dominant (dots) or suppressed (crosses). This plot illustrates that individual switching events (say, termination of A's dominance, crosses) occur primarily during simultaneous negative fluctuations in $n_A(t)$ and positive fluctuations in $n_B(t)$. Clouds of points are oval rather than circular, indicating that population A can become suppressed even if it experiences a positive fluctuation, as long as population B simultaneously receives a large enough transient positive input. *C*: comparison with Lankheet's (2006) experimental STAs. Timescale of the noise in the model was lengthened (500- instead of the 100-ms timescale used previously) to match the slowly filtered noisy inputs used in the experiment. STAs (solid lines) show sharp positive and negative peaks just before the transition, which are preceded by wider and shallower opposite peaks. The presence of the latter peaks was not due to the interaction between the timescale of the noise and the long timescale of the adaptation: the same simulation with no adaptation reproduced the shallow peaks (dashed lines; to keep the mean dominance durations at about 4 s, the strength of the noise was increased from $\sigma = 0.03$ to $\sigma = 0.05$).

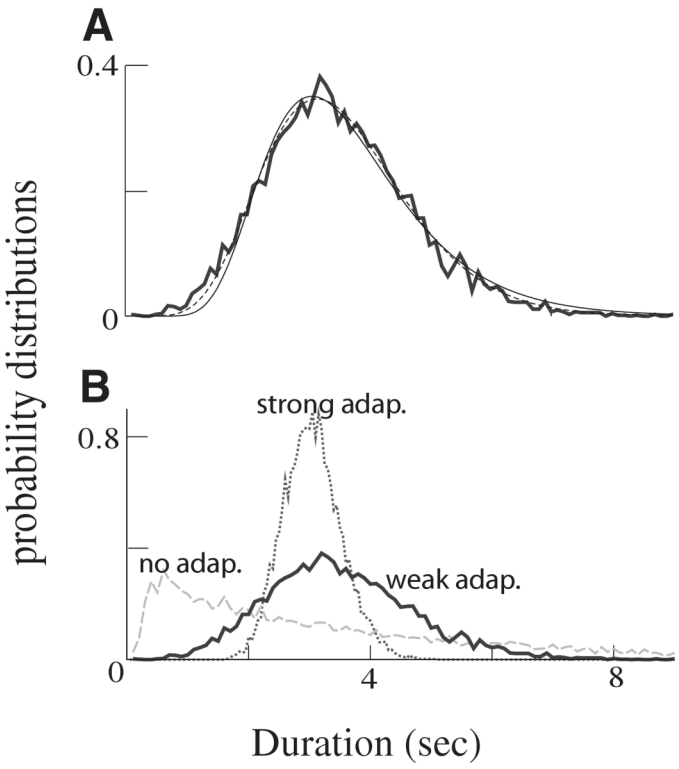

FIG. 7.

A combination of strong noise and weak adaptation provides a fit to experimental duration distributions. *A*: distribution of dominance durations resulting from the 2-variable rate-based model (thick line) is plotted along with log-normal (thin line) and gamma (dashed line) fits (see APPENDIX D for more details). *B*: distributions for 3 different amplitudes of adaptation, $\gamma = 0.2$ (strong), 0.1 (weak; same as in *A*), and 0 (no adaptation). Mean dominance period is kept approximately constant in all cases by setting $\sigma = 0.0125$, 0.03, and 0.04, respectively. In all cases $g_A = g_B = 0.01$.

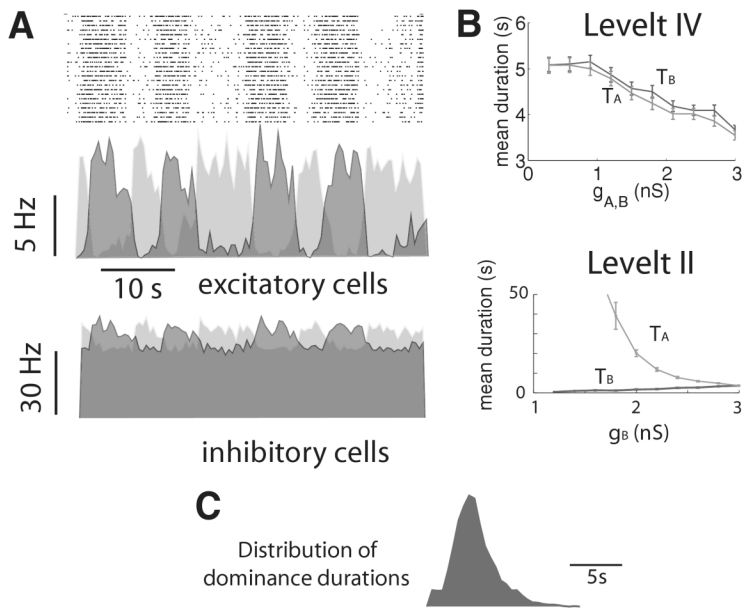

FIG. 8.

Rivalry alternations in the spiking neural network. *A, top*: raster plot of the spike times of neurons belonging to an excitatory population. *Middle*: mean firing rates of the 2 competing excitatory populations show alternations. *Bottom*: mean firing rates of the 2 corresponding inhibitory subpopulations. In all 3 panels, inputs are the same for the 2 populations: $g_{A,B}$ = 2 nS. *B*: spiking network behaves in accordance with Levelt's propositions IV (*top*) and II (*bottom;* here g_A is fixed at 3 nS). *C*: distribution of dominance durations has a skewed Gaussian shape (inputs as in *A*).

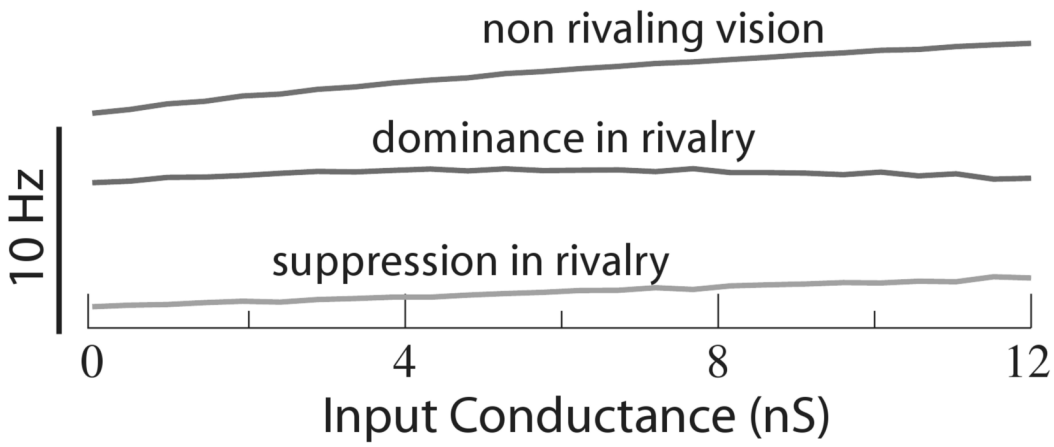

FIG. 9.

Activity during rivalry is lower than that during nonrivaling vision. Mean population firing rate during nonrivaling vision, and in the dominant and the suppressed states during rivalry, as a function of stimulation strength (equal to both populations in the rivalry case and zero to the competing population in the nonrivalry case).

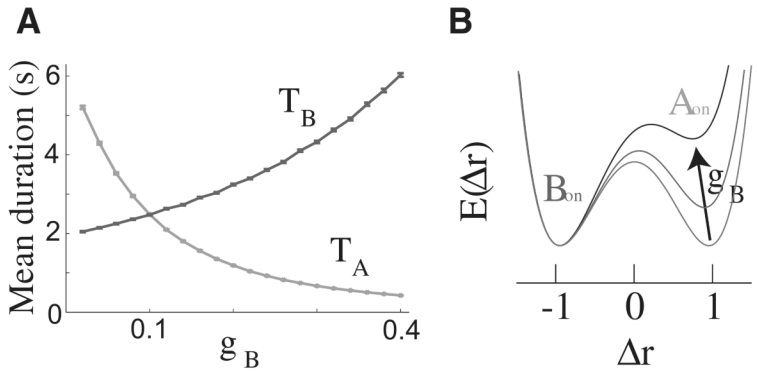

FIG. 10.

Dominance durations at large input strengths. *A*: mean dominance durations for the energy model in Eq. 2 as a function of g_B for fixed $g_A = 0.1$. When g_B becomes larger than g_A, the mean dominance duration for percept *B* is mostly affected, unlike what happens in the range $g_B < g_A$ (see also Fig. 2B) and in violation of Levelt's (1968) proposition II. *B*: energy function, Eq. 1, for several values of $g_B \geq g_A$ ($g_B = 0.1$, 0.2, and 0.4); g_A is fixed at 0.1. Percept *A* loses stability and percept *B* gains stability.