

Swarthmore College

Works

Mathematics & Statistics Faculty Works

Mathematics & Statistics

2023

New Insights Into Binocular Rivalry From The Reconstruction Of Evolving Percepts Using Model Network Dynamics

Kenneth Barkdoll , '24

Yuhua Lu , '24

Victor J. Barranca

Swarthmore College, vbarran1@swarthmore.edu

Follow this and additional works at: <https://works.swarthmore.edu/fac-math-stat>



Part of the [Mathematics Commons](#)

[Let us know how access to these works benefits you](#)

Recommended Citation

Kenneth Barkdoll , '24; Yuhua Lu , '24; and Victor J. Barranca. (2023). "New Insights Into Binocular Rivalry From The Reconstruction Of Evolving Percepts Using Model Network Dynamics". *Frontiers In Computational Neuroscience*. Volume 17, DOI: 10.3389/fncom.2023.1137015
<https://works.swarthmore.edu/fac-math-stat/267>



This work is licensed under a [Creative Commons Attribution 4.0 International License](#).

This work is brought to you for free by Swarthmore College Libraries' Works. It has been accepted for inclusion in Mathematics & Statistics Faculty Works by an authorized administrator of Works. For more information, please contact myworks@swarthmore.edu.



OPEN ACCESS

EDITED BY
Maurizio Mattia,
Italian National Institute of Health (ISS), Italy

REVIEWED BY
Hugh R. Wilson,
York University, Canada
Rodica Curtu,
The University of Iowa, United States
Alexander Pastukhov,
University of Bamberg, Germany

*CORRESPONDENCE
Victor J. Barranca
✉ vbarran1@swarthmore.edu

RECEIVED 03 January 2023
ACCEPTED 07 March 2023
PUBLISHED 24 March 2023

CITATION
Barkdoll K, Lu Y and Barranca VJ (2023) New insights into binocular rivalry from the reconstruction of evolving percepts using model network dynamics.
Front. Comput. Neurosci. 17:1137015.
doi: 10.3389/fncom.2023.1137015

COPYRIGHT
© 2023 Barkdoll, Lu and Barranca. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](#). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

New insights into binocular rivalry from the reconstruction of evolving percepts using model network dynamics

Kenneth Barkdoll, Yuhua Lu and Victor J. Barranca*

Department of Mathematics and Statistics, Swarthmore College, Swarthmore, PA, United States

When the two eyes are presented with highly distinct stimuli, the resulting visual percept generally switches every few seconds between the two monocular images in an irregular fashion, giving rise to a phenomenon known as binocular rivalry. While a host of theoretical studies have explored potential mechanisms for binocular rivalry in the context of evoked model dynamics in response to simple stimuli, here we investigate binocular rivalry directly through complex stimulus reconstructions based on the activity of a two-layer neuronal network model with competing downstream pools driven by disparate monocular stimuli composed of image pixels. To estimate the dynamic percept, we derive a linear input-output mapping rooted in the non-linear network dynamics and iteratively apply compressive sensing techniques for signal recovery. Utilizing a dominance metric, we are able to identify when percept alternations occur and use data collected during each dominance period to generate a sequence of percept reconstructions. We show that despite the approximate nature of the input-output mapping and the significant reduction in neurons downstream relative to stimulus pixels, the dominant monocular image is well-encoded in the network dynamics and improvements are garnered when realistic spatial receptive field structure is incorporated into the feedforward connectivity. Our model demonstrates gamma-distributed dominance durations and well obeys Levelt's four laws for how dominance durations change with stimulus strength, agreeing with key recurring experimental observations often used to benchmark rivalry models. In light of evidence that individuals with autism exhibit relatively slow percept switching in binocular rivalry, we corroborate the ubiquitous hypothesis that autism manifests from reduced inhibition in the brain by systematically probing our model alternation rate across choices of inhibition strength. We exhibit sufficient conditions for producing binocular rivalry in the context of natural scene stimuli, opening a clearer window into the dynamic brain computations that vary with the generated percept and a potential path toward further understanding neurological disorders.

KEYWORDS

neuronal networks, binocular rivalry, stimulus encoding, compressive sensing, non-linear dynamics, input-output mapping, autism

1. Introduction

A rich history of experiments demonstrate that when dissimilar constant visual stimuli are presented to each eye, humans typically perceive one of the monocular stimuli for a small period of time and then the other, with stochastic transitions between the two percepts as time advances (Dutour, 1760; Wheatstone, 1838). This phenomenon, referred to as *binocular*

rivalry, has garnered numerous studies exploring its intricacies and emergence, from both psychophysics and neuroscience perspectives (Breese, 1909; Blake and Fox, 1974; Leopold and Logothetis, 1996; Polonsky et al., 2000; Tong and Engel, 2001; Wilke et al., 2003). A diverse range of explanations have been hypothesized (Levelt, 1965; Tong, 2001; Blake and Logothetis, 2002), with various mathematical models probing the efficacy of different proposed mechanisms (Matsuoka, 1984; Dayan, 1998; Wilson, 2003; Said and Heeger, 2013). Collectively, these studies have primarily underlined internally-generated noise, a slow adaptation mechanism, and cross-column inhibition as potentially pivotal, though a completely parsimonious framework for the irregular percept switching and underlying stimulus encoding is still an active interdisciplinary area of inquiry (Kovacs et al., 1996; Logothetis et al., 1996; Tong et al., 2006; Sterzer et al., 2009).

Since the percept evolves while the monocular inputs remain constant, binocular rivalry offers unique insight into the structure-function relationship in the visual system. Binocular rivalry also provides an informative vantage point for studying internal computations in the brain, and it has been broadly applied in studying consciousness as well as cognition (Wilson, 2013; Koch et al., 2016; Vattikuti et al., 2016). Perceptual alternations are well-documented across many non-human mammals (Sengpiel et al., 1994; Carter et al., 2020) and outside the domain of vision as well, including in olfaction (Zhou and Chen, 2009), audition (Hupe et al., 2008), and touch (Holcombe and Seizova-Cajic, 2008), suggesting perceptual multistability results from relatively universal organizing principles.

Theoretical investigations thus far have often well-replicated key experimental observations, but they have either utilized idealized rate models (Shapiro et al., 2009; Li et al., 2017) or relatively simple stimuli (Laing and Chow, 2002; Cohen et al., 2019; Wang et al., 2020), such as those distinguished by one parameter, akin to the orientation of a grating. These phenomenological models largely do not seek to estimate the stimulus information encoded by the network dynamics, rather focusing instead on winner-take-all behavior and multistability properties to probe the neural substrates of perceptual inference in binocular rivalry (Gershman et al., 2012). Considering that sequentially alternating percepts are relatively rarely encountered outside of laboratory settings yet robustly occur for disparate complex stimuli in experiments (Baker and Graf, 2009; Miller, 2013), with dominance characteristics often modulated by higher order image features (Alais and Melcher, 2007), it is of significant interest to address the retention of complex monocular stimuli in the network dynamics and the possible role of network structure in rivalry as well as potential information loss.

In this work, we provide a new verifiable framework for characterizing the interplay between visual system connectivity, non-linear network dynamics, and stimulus encoding in binocular rivalry, utilizing a multi-layer neuronal network model with two competing ocular dominance columns driven by disparate realistic monocular image inputs. This study is among the first to explicitly use stimulus reconstructions based on spiking neuronal network activity to investigate binocular rivalry and provides a direct method of testing theories regarding its neural substrates. Our two-layer mechanistic network model reflects the effective connectivity between the retina and V1 (Barlow, 1981; Spear et al., 1996).

In the second layer, we introduce two pools of neurons with individually balanced dynamics driven by different images, cross-column competition *via* long-range connections from excitatory to inhibitory neurons, and spike-frequency adaptation to account for the common mechanisms believed to together be necessary for binocular rivalry. Each of these key assumptions is addressed in turn in this introduction.

We assume that the two downstream pools compete for dominance and each corresponds to a different attractor, such that when the network-averaged firing rate of one pool is significantly larger than that of the other pool, the percept corresponding to the more active pool is selected. Systems demonstrating competition among clusters of nodes display useful computational characteristics beyond decision making, such as sequence learning (McKinstry et al., 2016), classification (Krizhevsky et al., 2017), and signal restoration (Rutishauser et al., 2011). Competition in our model network stems from mutual inhibition between the pools that is facilitated by long-range connections from excitatory neurons in one pool to inhibitory neurons in the other, agreeing with experimental evidence that such long-range connections likely originate from excitatory neurons (Stettler et al., 2002; Douglas and Martin, 2004; Tamamaki and Tomioka, 2010; Binas et al., 2014).

Without slow adaptation or noise, a single percept may be selected for all time and the system would thus display winner-take-all behavior. In phenomenological models of binocular rivalry, adaptation is commonly implemented by either synaptic depression, in which neurotransmitters or vesicles are depleted over time, or spike-frequency adaptation (Freeman, 2005; Kilpatrick and Bressloff, 2010). There are two main slow currents that promote spike-frequency adaptation through gradual negative feedback, namely the non-inactivating muscarinic potassium current and the after-hyperpolarization (AHP) current, which together largely determine the firing threshold and slope of the neuronal voltage trace (Yamada et al., 1989; Ermentrout et al., 2001). Since the two main adaptation mechanisms both successfully decrease the excitability of the dominant pool after continued firing activity to help facilitate a switch in percept and there is no clear consensus regarding which mechanism is most closely implicated in binocular rivalry, we incorporate spike-frequency adaptation into our model for concreteness.

To account for the irregular nature of the percept switching in binocular rivalry, we assume each pool in the second layer demonstrates balanced dynamics in isolation (van Vreeswijk and Sompolinsky, 1996; Troyer and Miller, 1997), such that fluctuations in activity are internally generated. While here we assume irregular dynamics are the result of a chaotic attractor, noise is often instead explicitly incorporated into models of binocular rivalry and serves an analogous role in breaking symmetry to facilitate a new percept. Though either mechanism can provide an effective contribution to the stochastic switching, our choice of balanced dynamics reflects compelling experimental evidence that strong excitatory and inhibitory inputs into a given neuron are dynamically counteracted in time such that input fluctuations result in irregular spiking activity (Haider et al., 2006; Miura et al., 2007; London et al., 2010). Such balanced dynamics are theorized to have diverse functional benefits, facilitating robust spatial working memory (Lim and Goldman, 2014), effective predictive coding (Boerlin et al., 2013),

efficient food source selection for honeybees (Borofsky et al., 2020), and successful mammalian learning (Fiete et al., 2007; Ingrassio and Abbott, 2019).

While our model exhibits highly non-linear and irregular dynamics with complex network structure, we nonetheless use coarse-graining techniques to derive an approximate linear input-output mapping. Linking the pixels composing the image stimuli to the evoked network dynamics, this map furnishes the reconstruction of the evolving dominant monocular image inputs and demonstrates the successful retention of stimulus information during binocular rivalry. We utilize a dominance metric based on the firing rates of the two downstream pools to identify the dominant pool during each period of dominance, exhibiting persistent bistability and irregular switching in the model dynamics. Using only the time-averaged activity data collected during a physiologically realistic dominance period, we are able to generate accurate dynamic percept reconstructions for both simple gratings as well as natural scene monocular inputs. The efficacy of these reconstructions is further improved when more biologically realistic spatial receptive fields are incorporated into the feedforward connectivity for the two pools, agreeing with the evolutionary selection of this network structure.

It is important to note that these reconstructions succeed despite our assumption that the monocular input images contain significantly more pixels than downstream neurons in the two pools, reflecting the compressive nature of the early visual system that is typically unaccounted for in model-based studies of binocular rivalry (Barlow, 1981; Barranca et al., 2014b). In the human visual system, for example, 150 million photoreceptors are processed downstream by only about 1.5 million retinal ganglion cells. Further downstream, information encoded by the millions of neurons in the lateral geniculate nucleus (LGN) is later expanded in the primary visual cortex (V1), which is composed of roughly 40 times as many neurons as the LGN (Spear et al., 1996). Compressive pathways are exhibited in most other sensory systems (Welker, 1976; Knudsen and Konishi, 1978; Mori et al., 1999; Brecht and Sakmann, 2002), and are hypothesized to play a central role in efficient coding in the brain (Barlow, 1961). Reflecting information loss through the sensory bottleneck and the initial time necessary for a monocular stimulus to dominate, our model demonstrates the experimentally documented transient period in the first few hundred milliseconds of binocular rivalry, resulting in heavily diminished percept reconstruction quality and likely reflecting the superposition of the monocular stimuli in the percept often reported during this period (Blake et al., 1991; Pearson and Brascamp, 2008).

By methodically varying the strengths of the monocular stimuli presented to the two pools, we establish agreement with Levelt's four laws, which commonly serve as a benchmark for theoretical investigations of perceptual bistability due to their ubiquity across years of experiments (Levelt, 1965). We additionally show that the dominance durations in our model are right-skewed and gamma distributed, agreeing with standard experimental observations (Kovacs et al., 1996). In exploring perturbations to the model dominance durations, we provide further insight into the precise mechanisms for the percept switching and how rivalry dynamics crucially depend on stimulus strength.

Finally, we utilize the flexible nature of our model network connectivity and our ability to directly probe its parameters to better understand neurological disorders. There is evidence that individuals with autism demonstrate a relatively slow rate of percept alternations in binocular rivalry (Robertson et al., 2013; Spiegel et al., 2019), and experiments also indicate that an imbalance in inhibitory and excitatory neuronal inputs may underlie autism spectrum disorders (Gao and Penzes, 2015; Nelson and Valakh, 2015; Rosenberg et al., 2015). Based on these two findings, we systematically scale down the strength of inhibition in our model, reflecting the potentially reduced inhibition in the autistic brain, and consequently produce increasingly long dominance durations. Providing new credence to the excitation/inhibition hypothesis for autism, this result strengthens the possibility of restoring balanced inhibition and excitation as a potential avenue for treating autism.

2. Results

2.1. Mathematical model of binocular rivalry with direct monocular image drive

The two-layer network model we consider has two competing pools of downstream neurons that are each driven by distinct monocular image inputs, as represented by the schematic in Figure 1A. In light of several lines of experimental evidence indicating a link between percepts reported in binocular rivalry and the activity in both V1 (Logothetis et al., 1996; Polonsky et al., 2000) and other parts of the visual system (Haynes et al., 2005), we view the feedforward architecture of our model as a phenomenological reflection of the effective connectivity between photoreceptor output in the retina and V1, abstracting over the detailed connectivity structure and diverse neuron types in between these areas (Field and Chichilnisky, 2007; Anderson et al., 2011). Since it may potentially be the case that a hierarchy of several network layers are all partially responsible for various features of binocular rivalry (Wilson, 2003), we aim to include more general characteristics in our network connectivity and focus on dynamic percept encoding in the context of the key mechanistic contributing factors.

Each downstream pool is composed of N neurons, such that N_E are excitatory and N_I are inhibitory, with each neuron exhibiting pulse-coupled integrate-and-fire (I&F) dynamics. Though relatively idealized, the I&F model has been shown to well replicate experimentally observed subthreshold voltage dynamics and has been utilized to great effect in numerous large-scale model-based studies (Abbott, 1999; Rauch et al., 2003; Burkitt, 2006; Rangan and Cai, 2006; Mather et al., 2009; Barranca et al., 2014a). In formulating the detailed recurrent network connectivity and resultant dynamical regime, we utilize a balanced network configuration within each pool (van Vreeswijk and Sompolinsky, 1996; Miura et al., 2007; Barranca and Zhou, 2019; Barranca et al., 2019), producing the asynchronous dynamics in the dominant pool that underlies the stochastic switching phenomenon and our derivation of the network input-output transformation. While there may be other plausible noise-based mechanisms for the variability observed in dominance durations (Freeman, 2005),

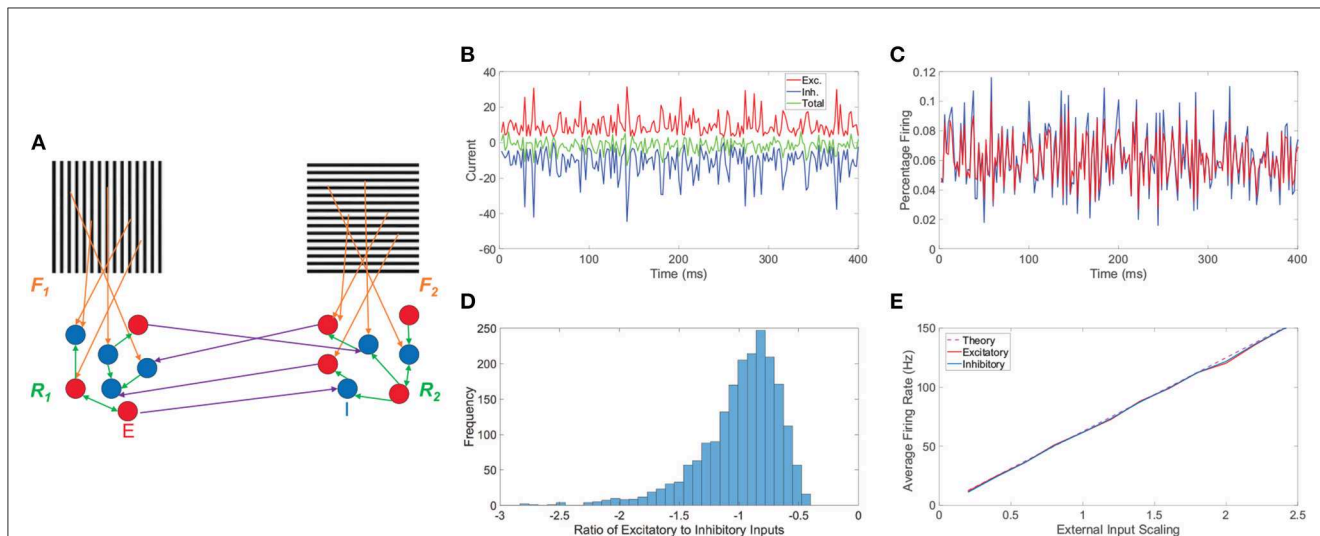


FIGURE 1 Network model and balanced dynamics. **(A)** Two-layer network with competing downstream pools schematic. Pool 1 and pool 2 are composed of integrate-and-fire neuronal networks driven by distinct monocular stimuli via feedforward connectivity matrices, F_1 and F_2 , respectively (orange). Recurrent interactions between neurons within pool 1 and pool 2 are given by matrices R_1 and R_2 , respectively (green). Long-range connections between neurons in the two pools (purple) originate at excitatory neurons (red, E) and terminate at inhibitory neurons (blue, I). **(B)** Time-evolving excitatory (red), inhibitory (blue), and total (green) inputs into a sample excitatory neuron in a balanced network. **(C)** The corresponding percentage of excitatory neurons firing (red) and inhibitory neurons firing (blue) as a function of time. **(D)** Histogram of the ratio between total excitatory input and total inhibitory input across neurons in the balanced network. **(E)** Gain curves depicting the population-averaged firing rates across a pool of downstream neurons as a function of the external input strength scaling. The excitatory population is plotted in red, the inhibitory population in blue, and the theoretical firing rate from Equation (8) in dashed purple. Note parameters are chosen such that theoretically the excitatory and inhibitory populations should have the same average firing rate. Panels **(B–E)** consider the balanced network dynamics of an isolated single downstream pool in the model prescribed by schematic **(A)** in the absence of both long-range connections between pools and spike-frequency adaptation. Parameters utilized are $R_{EE} = R_{IE} = 1, R_{II} = -1.8, R_{EI} = -2, N_E = 1000, N_I = 1000, f_E = 1, f_I = 0.8, m_0 = 0.5, K = 40, \theta_E = 1,$ and $\theta_I = 0.8$.

balanced networks have recently been incorporated into the study of binocular rivalry (Cohen et al., 2019; Wang et al., 2020) and largely produce dynamics that agree with core experimental findings on both the distribution of dominance durations and the variability of neuronal activity more broadly.

The voltage (membrane-potential) of the i^{th} neuron in the k^{th} population of the first downstream pool in the network, v_{1k}^i , has activity dictated by the dynamical system below (subscripts $k = E$ and $k = I$ denote excitatory and inhibitory neurons, respectively)

$$\begin{aligned} \frac{dv_{1k}^i}{dt} = & -g_L(v_{1k}^i - V^{Re}) + \sum_{\substack{j=1 \\ j \neq i}}^{N_E} R_{1kE}^{ij} \sum_l \delta(t - \tau_{1l}^{jl}) \\ & + \sum_{\substack{j=1 \\ j \neq i}}^{N_I} R_{1kI}^{ij} \sum_l \delta(t - \tau_{1l}^{jl}) + \sum_{\substack{j=1 \\ j \neq i}}^{N_E} C_{2kE}^{ij} \sum_l \delta(t - \tau_{2l}^{jl}) \\ & + \sum_j F_{1k}^{ij} p_1^j, \end{aligned} \tag{1a}$$

$$\frac{d\theta_{1k}^i}{dt} = -\lambda(\theta_{1k}^i - \theta_k), \tag{1b}$$

evolving from reset potential, V^{Re} , until increasing enough to reach its firing threshold, θ_{1k}^i , at which time the neuron spikes. Once a neuron fires, its voltage is instantaneously reset to the value V^{Re} and the voltages of all post-connected neurons are adjusted upon integrating over the Dirac delta functions $\delta(\cdot)$ in Equation (1a). The spike times for the i^{th} neuron in the k^{th} population of the first

pool are denoted τ_{1k}^{il} and are indexed by $l = 1, 2, \dots$ in ordering the firing events in time. Note that the voltage dynamics of the neurons in the second downstream pool are analogously defined except the locations of the initial subscripts 1 and 2 in Equation (1) are interchanged.

The $N \times N$ recurrent connectivity matrix for the first downstream pool, \mathbf{R}_1 , determines the instantaneous neuronal interactions within the pool upon firing events and is indexed such that R_{1kl}^{ij} denotes the recurrent connection strength between the i^{th} post-connected neuron in the k^{th} population and the j^{th} pre-connected neuron in the l^{th} population. Each entry of the recurrent connectivity matrix is prescribed by a Bernoulli distribution, where $R_{1kl}^{ij} = R_{kl}/\sqrt{K}$ with probability K/N_I and $R_{1kl}^{ij} = 0$ otherwise. Thus, on average, a given downstream neuron will be post-connected to K excitatory and K inhibitory neurons within its pool, where the excitatory connection strength R_{kE} is positive and the inhibitory connection strength R_{kI} is negative. The recurrent connectivity is assumed sparse in this model, as typically observed in experiment (He et al., 2007; Ganmor et al., 2011), implying $1 \ll K \ll N_E, N_I$.

Each downstream pool corresponds to a column of neurons with a specific ocular dominance preference, which is well-documented in V1 (Blasdel, 1992; Issa et al., 1999). Reflecting ocular preference, a neuron in a given downstream pool receives external drive determined by its corresponding monocular stimulus, which is mediated by a feedforward connectivity matrix. In the first pool, for example, the corresponding monocular stimulus is given by a constant image input vector \mathbf{p}_1 with

$n = 10N_E$ total pixel components, modeling the large number of photoreceptors relative to downstream neurons in V1 (also relative to the number of neurons in earlier layers, such as the retinal ganglion cells or neurons in the LGN). We have verified that a less extreme compression ratio, focused on connectivity between alternative specific layers, may be used with minimal impact on our analysis. The feedforward connectivity matrix, F_{1k} , determines the connections between the upstream and downstream layers, where F_{1k}^{ij} denotes the feedforward connection strength between the j^{th} component of the stimulus vector, p_1^j , and the i^{th} neuron in the k^{th} population of the first pool. Note, however, we do not consider neurons with orientation selectivity (Ferster et al., 1996; Tao et al., 2006) and we view the monocular image input as a single patch of the visual scene sampled by the downstream pool in V1 (Hubel and Wiesel, 1962) for analytical tractability and more direct mechanistic insights that apply to stimuli beyond simple oriented gratings. With the model network connectivity characterized, we now proceed to underline the three main model features that are together sufficient to produce perceptual alternations in binocular rivalry.

2.1.1. Irregular firing activity

Irregular firing activity is internally generated by the recurrent interactions among the downstream neurons, as we assume the external input vectors forcing the two pools are constant, such that the expected external drive into a downstream neuron in the k^{th} population of either pool is $\sqrt{K}f_k m_0$. Here, m_0 and f_k are $\mathcal{O}(1)$ parameters applying to both pools, where m_0 scales the overall external drive strength for the two populations and f_k further scales the external drive strength for the k^{th} population only. For analytical tractability in our initial investigation, we assume the feedforward connectivity matrices are sparse with entries determined by Bernoulli distributions akin to the recurrent connectivity matrices. Later, in Section 2.5, we instead consider the impact of utilizing more realistic spatially localized feedforward connectivity.

2.1.2. Competition

Competition between the two downstream pools is fostered by long-range connections from excitatory neurons in one pool to inhibitory neurons in the other pool. The long-range connections starting from neurons in the second pool are given by matrix C_2 , where C_{2kl}^{ij} denotes the long-range connection strength between the i^{th} post-connected neuron in the k^{th} population of the first pool and the j^{th} pre-connected neuron in the l^{th} population of the second pool. The entries of C_2 are prescribed similarly to those of the recurrent connectivity matrices except the only non-zero long-range connections possible are from excitatory to inhibitory neurons, so $C_{2IE}^{ij} = R_{IE}/\sqrt{K}$ with probability K/N_E and $C_{2kl}^{ij} = 0$ otherwise. In this case, when one pool is highly active, it will send numerous excitatory inter-pool impulses to inhibitory neurons in the other pool and thereby tends to suppress the activity of the competing neurons. Though a host of mechanisms for inter-pool competition exists, such long-range excitatory connections are a plausible and biologically feasible apparatus for inter-pool competition (Stettler et al., 2002; Douglas and Martin, 2004; Tamamaki and Tomioka, 2010; Binas et al., 2014), and they have

been included in prior realistic models of binocular rivalry (Wang et al., 2020).

In an isolated pool that receives no cross-pool connections, reminiscent of the dominant pool that receives only weak competitive drive from the suppressed pool, it is important to underline that if the recurrent connectivity strength parameters, R_{kl} , are $\mathcal{O}(1)$, then only $\mathcal{O}(\sqrt{K})$ excitatory impulses are needed for a constituent neuron to fire given an $\mathcal{O}(1)$ firing threshold. Hence, the mean excitatory and inhibitory inputs into each downstream neuron are each in total of the same order as the default firing threshold, reflecting the dynamic balance of large excitatory and inhibitory inputs observed in many experimental settings (Britten et al., 1993; Haider et al., 2006; Miura et al., 2007; London et al., 2010; Xue et al., 2014). As a result, intermittent fluctuations in neuronal input are largely responsible for firing events as well as their irregular distribution. Since the excitatory and inhibitory inputs into a given downstream neuron in the balanced state dynamically cancel over time, a nearly constant level of asynchronous activity is typically produced across neurons in the dominant pool (van Vreeswijk and Sompolinsky, 1996; Barranca et al., 2019).

2.1.3. Spike-frequency adaptation

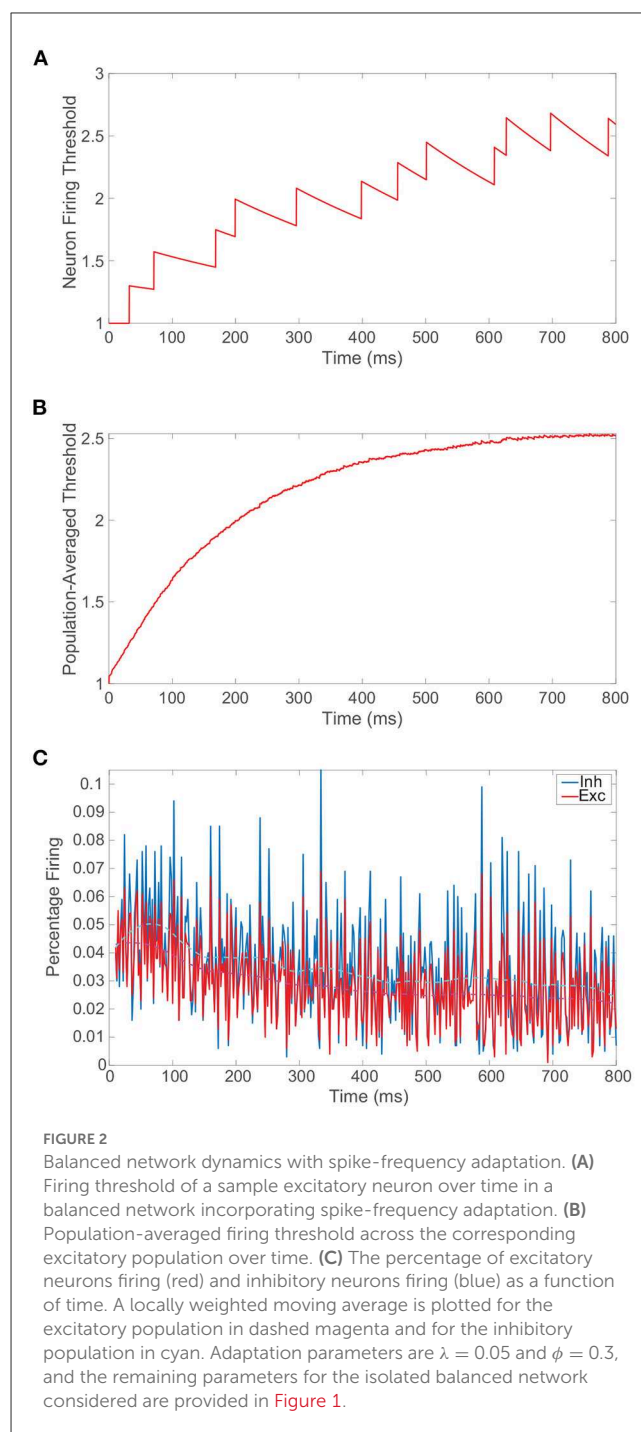
Spike-frequency adaptation is the final major feature incorporated into the model (Brown and Adams, 1980; Benda and Herz, 2003; Barranca et al., 2014a), supported by a large body of theoretical evidence that underlines the resultant decrease in neuronal firing rate over time in response to a constant stimulus as a core contributor to the percept switching in binocular rivalry. In our model, adaptation arises from a dynamic firing threshold for each neuron, rather than a static threshold as in traditional balanced networks. As time progresses, this dynamic firing threshold generally rises for the dominant pool and consequently increases the excitation needed for its neurons to undergo action potentials. Multiple subsequent spikes occurring close in time sum to produce an accumulated increase in firing threshold and ultimately reduce a neuron's firing frequency. In particular, the firing threshold of the i^{th} neuron in the k^{th} population of the first downstream pool, $\theta_{1k}^i(t)$, increases by a fixed positive constant ϕ at the moment that neuron fires. Between firing events, $\theta_{1k}^i(t)$ evolves according to Equation (1b), such that the dynamic firing threshold decays to the constant non-adapted firing threshold for all neurons in the k^{th} population, θ_k , in the absence of firing events, with speed dictated by decay rate constant λ . The voltages and thresholds are non-dimensionalized such that $V^{Re} = 0$, $\theta_E = 1$, and $\theta_I = 0.8$, with $g_L = 50\text{s}^{-1}$ corresponding to the standard membrane-potential time-scale of 20 ms (McLaughlin et al., 2000; Brette et al., 2007; Barranca et al., 2014b).

In the context of this study, adaptation further facilitates rivalry since it causes the activity of the dominant pool to weaken with time, giving the suppressed pool the opportunity to again dominate as the firing thresholds of its neurons decrease and consequently the overall firing activity of the suppressed pool increases with time. Without relatively irregular neuronal dynamics, such alternations in dominance could occur almost periodically with highly correlated dominance durations. In

order to agree with the gamma-distributed dominance durations typical in experimental data (Kovacs et al., 1996), irregularity is intrinsically introduced through balanced network structure, and prior theoretical work suggests that binocular rivalry may only occur with the appropriate mix of irregularity and adaptation (Shapiro et al., 2009).

Before turning to the rivalrous behavior of the full model, we provide useful insight into the model network dynamics by underlining several main characteristics of balanced activity for a single isolated pool in the absence of spike-frequency adaptation in Figure 1. For a sample downstream neuron, we plot in Figure 1B its total excitatory and inhibitory inputs, which dynamically cancel over time and are much larger in magnitude than the static firing threshold, causing the neuronal voltage to only sporadically reach the firing threshold. As a result, on the population scale, we observe in Figure 1C a nearly constant level of asynchronous firing activity with only a relatively small percentage of neurons spiking at any point in time. For a given neuron, the time-averaged ratio between its total excitatory and total inhibitory input is primarily near -1 . Since this quotient is always negative when well-defined and a ratio larger in magnitude than -1 implies an excess of excitatory input, excitatory to inhibitory input ratios near -1 demonstrate that on a neuron-by-neuron basis the excitatory and inhibitory inputs are largely proportional. A histogram of these ratios across the pool of neurons is given in Figure 1D, showing a clear center near -1 and thus widespread balance. Despite the highly non-linear voltage dynamics of neurons on an individual level, the network firing rate response properties of the isolated balanced network are quite linear. In Figure 1E, we adjust the external input by increasing the overall scaling strength m_0 and depict the resultant population-averaged firing rates. We observe that as the mean external drive is increased, the population-averaged firing rate of both the excitatory and inhibitory populations linearly amplifies for this finite network realization and agrees with the theoretical gain curve derived in the large-network limit in Section 4.1.

We conclude the preliminary discussion of our model network by examining the impact of the spike-frequency adaptation prescribed by Equation (1b). In Figure 2A, we plot the dynamics of the firing threshold for a sample neuron subject to moderately strong adaptation, demonstrating relatively large upward jumps in threshold when that neuron fires and a slow decay in the dynamic threshold between its spikes. The neuronal firing threshold shows a more rapid overall increase initially and then the threshold ultimately saturates at an elevated level once the neuron tends to spike more slowly. On a more global level, in Figure 2B we depict the population-averaged threshold dynamics with time for the excitatory neurons, which, albeit more smoothly, also exhibit a relatively rapid initial ascent and saturation in the long-time limit. Consequently, as seen in Figure 2C, the percentage of neurons firing decreases initially in time and then remains nearly constant with minor fluctuations about a now depressed mean. As demonstrated in prior work (Barranca et al., 2019), so long as the adaptation is not too strong, the network still demonstrates balanced dynamics with lower overall activity following an initial transient period. When spike-frequency adaptation is included only in the excitatory neurons, the parameter regime in which balanced dynamics are theoretically expected in fact broadens.



2.2. Binocular rivalry dynamics and model percept reconstructions

In order to reconstruct the dynamic percept during binocular rivalry, we derive a generalized input-output map capable of incorporating multiple stimuli in Section 4.1 and then construct a mechanism for selecting the appropriate percept at any given point in time. While previous work has developed such mappings in balanced networks using coarse-graining techniques rooted in statistical mechanics for I&F neurons without adaptation

(Barranca, 2022) or binary neurons with adaptation (Barranca et al., 2019), the model in this work is more detailed and requires an extension of these methodologies. We note, however, that even if the network parameters are not strictly in the theoretical balanced operating regime, as long as the evoked dynamics are sufficiently irregular, our derivation of the input-output mapping still holds, making our framework robust to alternative model settings.

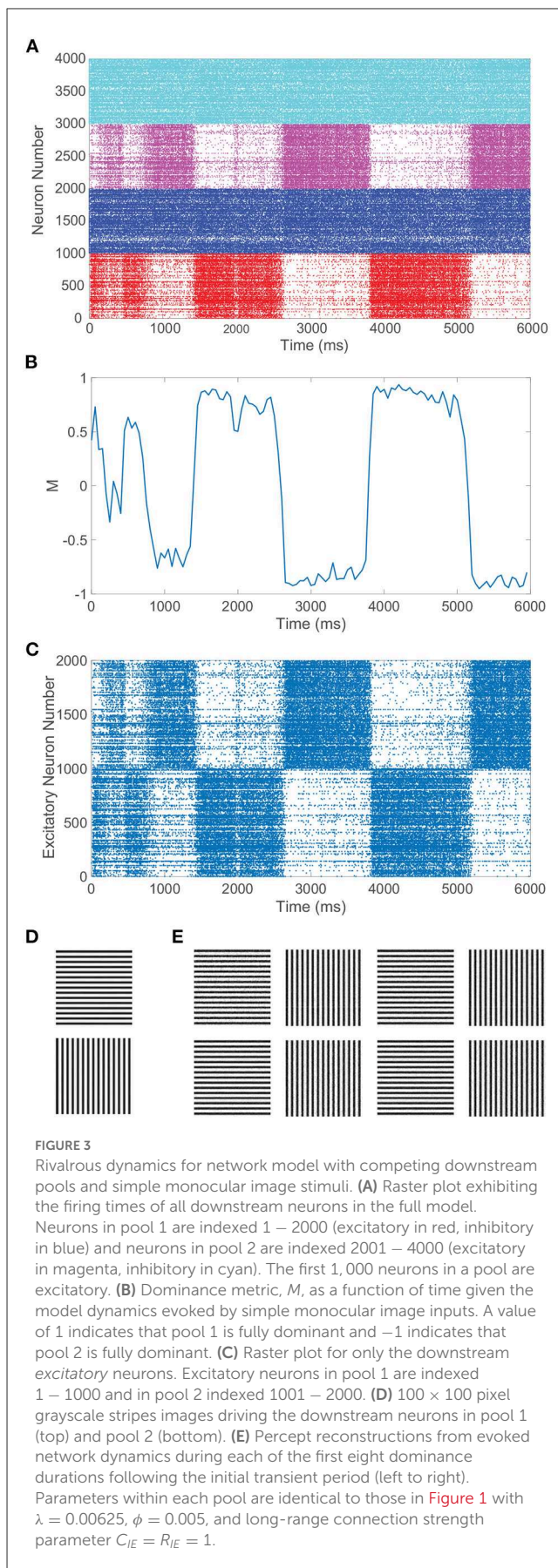
A typical requirement for balanced dynamics is that activity in both the excitatory and inhibitory populations remains non-zero and asynchronous in the large-network limit. Thus, we require the population-averaged firing rates in a given pool, m_{zk} for $k = E, I$ and $z = 1, 2$, to obey $0 < m_{zk} < \infty$ as $N \rightarrow \infty$ and as $K \rightarrow \infty$ given a fixed ratio N_E/N_I . In order to hold well for a finite-network realization, the linear input-output mapping and associated theoretical bounds on the model parameters require that the network is sufficiently large and composed of enough excitatory and inhibitory connections to generate sustained irregular activity.

In particular, we require balanced dynamics for neurons in the currently dominant pool, as excitatory neurons in the suppressed pool are expected to be significantly less active and more inhibited than assumed in a balanced network. Since the inter-pool connections are from excitatory to inhibitory neurons, the inhibitory neurons in the suppressed pool will typically continue firing from this additional cross-pool excitation while the excitatory neurons in the suppressed pool will primarily be quiescent. This is demonstrated in the raster plot in Figure 3A for a model network simulation of 6,000 ms, showing persistent firing across both inhibitory populations for all time and significantly diminished firing activity across the excitatory neurons in the suppressed pool for each period of dominance. A relatively low level of firing in the non-dominant pool is observed in some experiments, consistent with the model dynamics (Blake and Logothetis, 2002), though whether the neurons firing in the suppressed pool are largely inhibitory in experiments is not yet fully determined.

The key input-output mapping, relating the monocular stimulus driving a downstream neuron in the first pool, \mathbf{p}_1 , to the vectors of evoked neuronal firing rates and expected voltages across the downstream neurons in the first pool, \mathbf{m}_1 and $\bar{\mathbf{v}}_1$, respectively, may be expressed compactly across the network in matrix form as

$$\bar{\mathbf{v}}_1 = \mathbf{V}^{\text{Re}} + \frac{1}{g_L} \left(\mathbf{F}_1 \mathbf{p}_1 + \mathbf{R}_1 \mathbf{m}_1 + \mathbf{C}_2 \mathbf{m}_2 - \mathbf{m}_1 \left(\bar{\boldsymbol{\theta}}_1 - \mathbf{V}^{\text{Re}} \right) \right), \quad (2)$$

where \mathbf{R}_1 is ordered such that the first N_E rows (columns) correspond to the N_E excitatory neurons and the next N_I rows (columns) correspond to the N_I inhibitory neurons in pool 1; we assume the same ordering for the remaining terms in Equation (2). The time-averaged threshold of each neuron in the first downstream pool, $\bar{\boldsymbol{\theta}}_1$, is included in the input-output mapping to account for the impact of spike-frequency adaptation, and the influence of inter-pool connections is accounted for by $\mathbf{C}_2 \mathbf{m}_2$. Note that this mapping is valid when the first pool is dominant. The input-output mapping for downstream neurons in the second pool is similarly given by interchanging the numbered subscripts in Equation (2) and is analogously derived. This mapping assumes the periods of dominance are sufficiently long such that the dynamical statistics, \mathbf{m}_1 , $\bar{\mathbf{v}}_1$, and $\bar{\boldsymbol{\theta}}_1$, are robust and well approximate long-time averages, which we have verified to be empirically true. In



practice, statistics regarding the network dynamics are computed separately for each period of dominance and then separately used to reconstruct each corresponding dominant monocular stimulus.

Since the firing activity of the excitatory neurons is most significantly modulated across periods of dominance, the dominant stimulus at time t will be determined by the population-averaged firing rates of the excitatory neurons in first and second pools, $m_{1E}(t)$ and $m_{2E}(t)$, respectively, computed in a small time bin centered around time t . Based on these firing rates, we compute a time-varying dominance metric that prescribes both the dominant pool and the degree to which it currently dominates,

$$M(t) = \frac{m_{1E}(t) - m_{2E}(t)}{m_{1E}(t) + m_{2E}(t)}. \quad (3)$$

The dominance metric varies between -1 and 1 , where a value closer to 1 indicates stronger dominance of the first pool and a value closer to -1 indicates stronger dominance of the second pool. We assume a dominance period begins when $M(t)$ switches sign and then remains above 0.4 in magnitude for at least 100ms .

Figure 3B depicts the dominance metric as a function of time for the same network simulation previously considered in Figure 3A. Here, we observe an initial transient period in which neither stimulus is fully dominant. Then we subsequently see clear alternating periods of dominance for each pool, with metric values hovering near -1 or 1 , along with fast transitions in the dominance metric value at the end of each period of dominance. As will be described in more detail in the next section, these dominance durations agree well with experimental observations and demonstrate irregularity in length. Note that the dynamics of the dominance metric specifically correspond to the rivalrous alternations in excitatory neuron firing activity shown for this simulation in Figure 3C, determining the duration of each dominance period, the rapid transitions in dominance, and the degree to which a given pool is dominant. In this case, the respective monocular stimuli driving the neurons in the two downstream pools are simple grayscale horizontal and vertical gratings, depicted in Figure 3D.

With a means of identifying the dominant stimulus as well as an input-output mapping, what remains to be determined is a method for reconstructing the appropriate percept given measurements of network dynamics during a period of dominance. In order to apply Equation (2) and recover the percept, only the monocular stimuli are assumed unknown and all remaining terms must either be known or estimated. The connectivity matrices are known exactly following a particular network realization and fixed model parameters, g_L and V^{Re} , are known as well.

Due to the assumed compressive nature of the feedforward pathway, i.e., since there are many more monocular stimulus components (pixels) than downstream neurons, reconstructing the dominant stimulus during each dominance duration from our input-output mapping requires solving a highly underdetermined linear system with infinitely many solutions. Addressing this issue, to select the percept most closely associated with the true dominant monocular stimulus, we apply compressive sensing (CS) theory to linear system (2). CS theory is a modern mathematical breakthrough that provides a means of efficiently sampling and reconstructing signals which are sparse in an appropriately-chosen

domain (Candes et al., 2006; Donoho, 2006; Candes and Wakin, 2008), and it has promoted discoveries in diverse applications across physics, biology, and image processing (Lustig et al., 2007; Bobin et al., 2008; Dai et al., 2009; Herman and Strohmer, 2009; Berger et al., 2010; Gross et al., 2010; Wang et al., 2011; Noor et al., 2013; Emad and Milenkovic, 2014). Natural scenes and gratings, common as monocular inputs in binocular rivalry, are known to have sparse representations in frequency-based spaces (Field, 1994) and thus CS techniques are amenable to their reconstructions. As discussed in detail in Section 4.2, compressive sensing reconstructions generally require linear measurements of static sampled data (Candes et al., 2006; Donoho, 2006), but our model demonstrates non-linear dynamics in time. The derived input-output mapping therefore allows us to successfully overcome this conceptual obstacle by approximating a linear relationship between the monocular stimuli and limited observations of the evoked network dynamics, facilitating efficient reconstructions of the dominant stimuli.

We use the network model output over a given dominance duration to estimate the terms reflecting dynamical statistics in Equation (2) and then apply a CS recovery algorithm known as the orthogonal matching pursuit (Tropp and Gilbert, 2007) to reconstruct the dominant stimulus. For a given period of dominance, we record the firing rates, time-averaged voltages, and time-averaged thresholds across all downstream neurons to obtain estimates for \mathbf{m}_z , \bar{v}_z , and $\bar{\theta}_z$, respectively. When the metric abruptly switches sign and signals an alternation, we stop collecting output data and perform a reconstruction. In the recovery process when the first pool is dominant, the stimulus \mathbf{p}_1 is considered the unknown input in the input-output mapping given by Equation (2), and the remaining terms are either known model parameters or measured from model simulation. In this case, the feedforward connectivity matrix \mathbf{F}_1 plays the role of the CS measurement matrix, and the solution obtained via the orthogonal matching pursuit, $\mathbf{p}^{\text{recon}}$, corresponds to the percept and is considered optimal when it is close to the true dominant stimulus \mathbf{p}_1 . After this reconstruction, we begin to recollect new output data throughout the next dominance duration for the subsequent reconstruction; this process of reconstructing and recollecting data continues throughout the simulation, giving a sequence of estimates for the percepts during the rivalrous alternations.

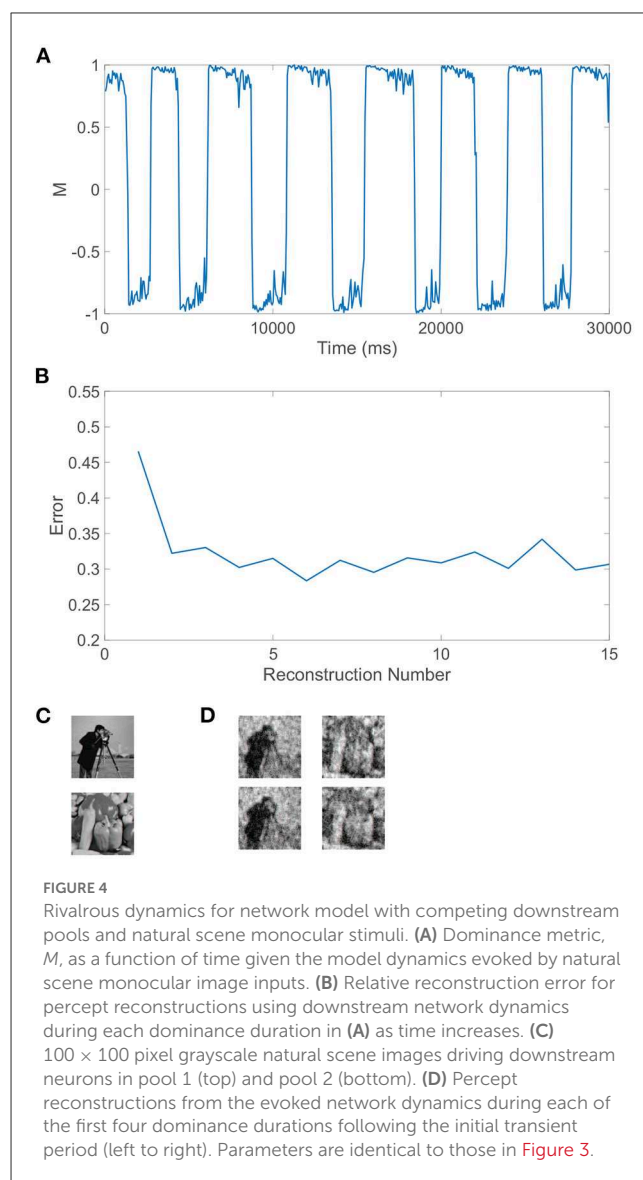
CS reconstructions of the switching percept using the network dynamics during each of the first eight dominance durations following the initial transient period are given in Figure 3E. In each case, the reconstruction is nearly perfect, showing successful encoding of the dominant percept in the downstream dynamics. We emphasize that there is minimal information loss from the non-linear dynamics and downstream compression for these relatively simple images. Note that the pixel values are normalized so that each monocular stimulus has the same average pixel value and the feedforward drive from the two input images is comparable. As a result, there is little systematic difference in both the dominance durations and the reconstruction quality corresponding to the two monocular stimuli.

It is important to emphasize that while it may be possible for the brain to implement CS recovery in general (Rozell et al., 2008), we make no claim that our precise method of reconstructing the

percept is performed by the visual system itself and hypothesize that summation implemented by a downstream binocular neuronal network or dichoptic differencing of monocular input *via* opponent neurons (Lansing, 1964; Bock et al., 2019) may be able to compute a dominance metric in the determination of the dynamic percept. Nonetheless, we see that rivalrous behavior can manifest even in the presence of potential information loss through a downstream sensory bottleneck and that monocular image information is well-encoded in the dynamics of the downstream neurons in the dominant pool. Even if network dynamics data is instead only collected during the first 200 ms of a given dominance period rather than throughout the entire dominance duration, we have verified that there is little change in the percept reconstruction quality and thus monocular stimulus information is indeed quite rapidly encoded.

We can similarly consider the model dynamics evoked when the two monocular stimuli are more complicated natural scenes, particularly those depicted in Figure 4C. In this case, we again observe clear alternations in the dominance metric plotted in Figure 4A, hovering near a value of 1 and then rapidly switching to gravitate near -1 , with such alternations continuing to occur irregularly throughout time. Each dominant monocular stimulus is analogously reconstructed *via* Equation (2) using data collected during each dominance duration. To measure the accuracy of a given stimulus reconstruction, $\mathbf{p}^{\text{recon}}$, we compute the relative reconstruction error, $\|\mathbf{p} - \mathbf{p}^{\text{recon}}\|/\|\mathbf{p}\|$, using the Euclidean norm, $\|\mathbf{p}\| = \sqrt{\sum_i p_i^2}$ and true dominant stimulus \mathbf{p} .

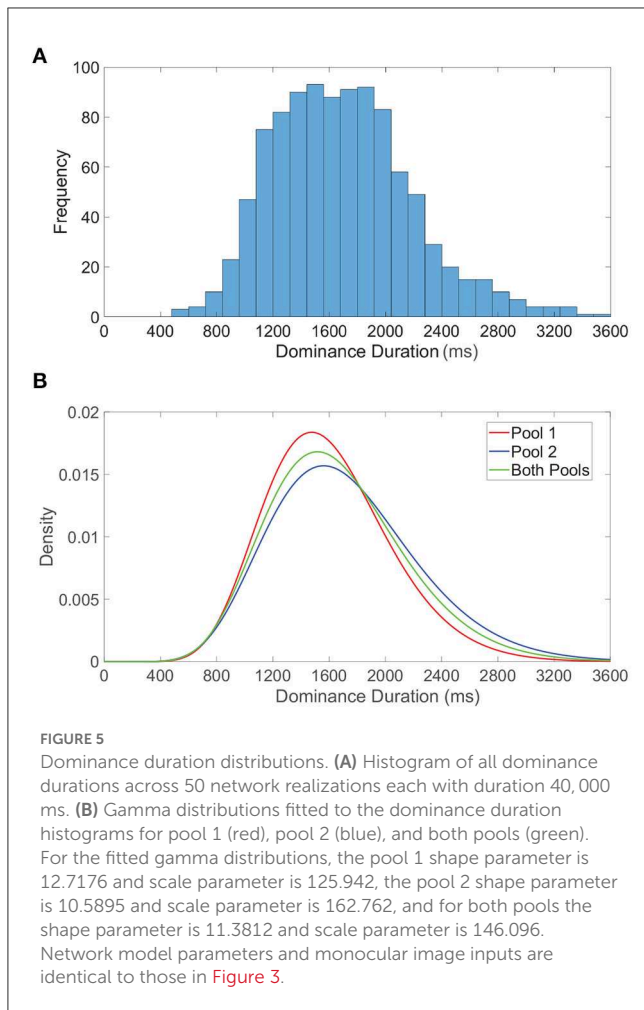
The relative error for each monocular natural scene reconstruction is plotted in Figure 4B over all dominance periods in the model simulation. Here, we include as the first data point the relative reconstruction error using network dynamics collected over the initial transient period, where the metric initially hovers slightly below 1 for a time shorter than a typical dominance duration. As expected from experimental results (Blake et al., 1991; Pearson and Brascamp, 2008), the reconstruction error in the transient period is relatively high and likely corresponds to the mixed percept often reported initially in binocular rivalry experiments. The relative reconstruction errors corresponding to each of the subsequent dominance periods are lower and display only minor fluctuations in value. We provide several sample percept reconstructions using data collected during each of the dominance durations following the initial transient period in Figure 4D. The reconstructions are each recognizable though less accurate than those obtained for the grating images considered previously. This is to be expected since these natural scenes display more complex structure in the sparse frequency space and yet only contain 10,000 pixels. For higher resolution images with more pixels and consequently more sparsity in the frequency domain, more accurate CS reconstructions are generally achievable given the same ratio of input image pixels to downstream neurons (Barranca et al., 2016). While high-frequency components are added when the resolution of an image is increased, the amplitude distribution of the dominant low-frequency components is often nearly identical to that of a lower resolution version of the image, and the additional downstream neurons introduced to maintain the same factor of compression will generally increase the accuracy of stimulus recovery since the nearly fixed dominant



frequency-components become better resolved when more neurons are available to sample them.

2.3. Dominance duration distribution and robustness to alternative architectures

Widespread experimental studies in both humans and non-human mammals indicate that dominance duration distributions in binocular rivalry are largely peaked, right-skewed, a few seconds in length, and conform well to a gamma distribution (Kovacs et al., 1996; Leopold and Logothetis, 1996). Due to their ubiquity and ease of application, such observations often serve as natural benchmarks for physiologically reasonable models of binocular rivalry. From the dynamics of the dominance metric displayed in Figures 3, 4, we see that model dominance durations are indeed often between 1 and 2 s with irregularity in their specific lengths.



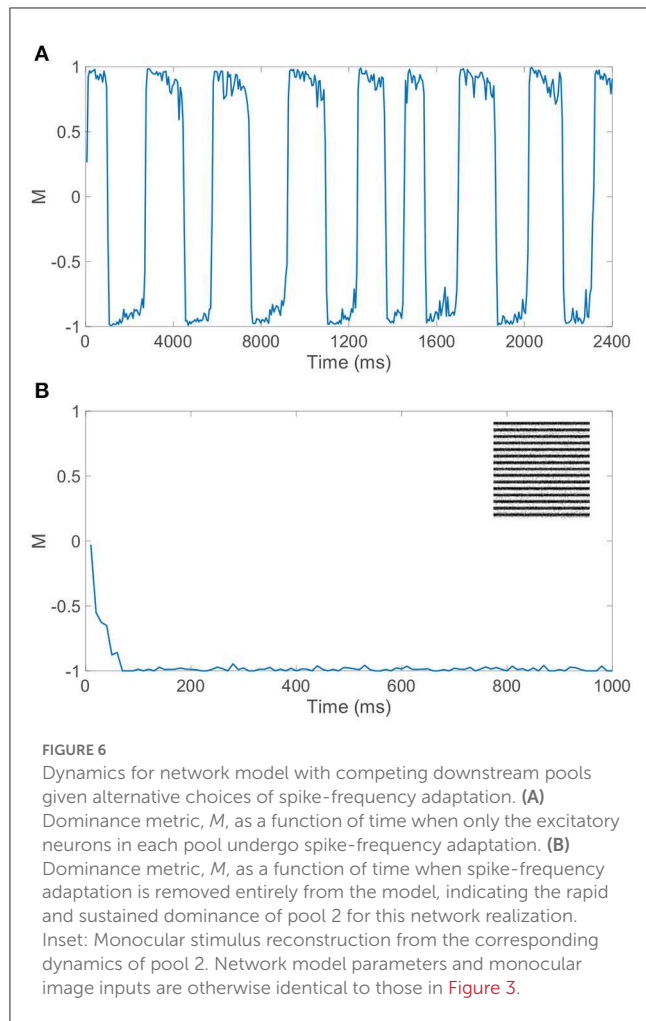
We examine the distribution of dominance durations in more detail in [Figure 5A](#), where we plot a histogram of all dominance durations for 50 model network realizations each with a total runtime of 40 s. This is meant to reflect the typical experimental setting in which rivalry dynamics are analyzed for multiple individuals (network realizations) for a relatively long time course. The resultant histogram is right-skewed with mean near 2 s, conforming to classical dominance duration statistics. In [Figure 5B](#), we similarly examine the dominance duration distributions for the first pool, second pool, and both pools, fitting each to a gamma distribution. Separately, the dominance durations in each pool are also well-approximated by a gamma distribution, with slight shifting in mean likely as a consequence of differences in feedforward connectivity and monocular input stimuli.

We note these peaked gamma distributions indicate that the probability of a subsequent switch immediately following an alternation is low, giving further credence to the notion that sufficient adaptation must accumulate, or alternatively wear off, before another percept switch can occur. It is important to emphasize that while dominance durations for a specific network realization are quite irregular, there is still more correlation in dominance durations for a particular network than across distinct network realizations. The inclusion of data accumulated over

multiple network realizations facilitates the especially long right tail in the dominance duration distributions, corresponding primarily to those network realizations that tend to produce especially long periods of dominance. This is consistent with the marked variations in average dominance duration across different individuals in experiments ([Gallagher and Arnold, 2014](#); [Bosten et al., 2015](#); [Bock et al., 2019](#)), and can potentially be attributed to underlying genetic or chemical differences ([Klink et al., 2010](#); [Miller et al., 2010](#); [Shannon et al., 2011](#)). It is worth noting that some studies report a correlation between sequential dominance durations in certain experimental settings ([van Ee, 2009](#)), especially demonstrating history dependence after prolonged exposure to a non-ambiguous display and highlighting the possible role of adaptation ([Nawrot and Blake, 1989](#)). While we do not explicitly consider how such correlations may be modulated or manifest, one approach to examining history dependence in the context of our model would be to include white noise in the dynamic firing threshold activity and examine its potential affect on serial correlations.

Gamma-distributed dominance durations and rivalrous switching are present for a relatively broad set of model parameters and ratios of excitatory to inhibitory neurons as long as the dynamics within the dominant pool are largely balanced. While the majority of our model simulations utilize a 1 : 1 ratio of excitatory to inhibitory neurons to ensure sufficiently well-balanced dynamics for a given finite network realization, we note that similar switching dynamics are observed when a 4 : 1 ratio of excitatory to inhibitory neurons is utilized consistent with estimates in V1 ([Gilbert, 1992](#); [Liu, 2004](#)). Moreover, if spike-frequency adaptation is only incorporated in excitatory neurons, holding the firing thresholds of all inhibitory neurons at θ_I , realistic rivalrous activity is also produced as depicted by the dominance metric dynamics in [Figure 6A](#). Experimental evidence indicates that excitatory neurons are generally more likely to undergo adaptation than inhibitory neurons ([La Camera et al., 2006](#); [Augustin et al., 2013](#)), and we see that binocular rivalry is consistent with this network setting as well. In fact, prior theoretical analysis demonstrates that the presence of spike-frequency adaptation only in the excitatory population can broaden the space of model parameters over which balanced dynamics manifest ([Barranca et al., 2019](#)).

If instead spike-frequency adaptation is removed from our model entirely, by fixing all excitatory neuron firing thresholds at θ_E and all inhibitory neuron firing thresholds at θ_I , rivalrous dynamics are not produced and there is a single dominant pool throughout time. As depicted in [Figure 6B](#), the dominance metric rapidly gravitates toward 1 or -1 , remaining near that value perpetually and failing to produce any switching even over a long time horizon. The identity of the dominant pool is random and dependent on the specific network realization, with balanced dynamics in the dominant pool as seen during a specific dominance period in the full model. This underlines the key role of a fatigue mechanism in binocular rivalry, and without irregularity in the model dynamics, which is produced under the balanced network paradigm in this case, the ubiquitous skewed dominance duration distribution is eliminated. Thus, we provide evidence for the necessity of both irregular network dynamics and slow adaptation, beyond the competition fostered by long-range connections between the downstream pools, to fully produce the stochastically switching percept that is the cornerstone of binocular rivalry.



2.4. Levelt's laws and the impact of stimulus strength on alternations

Levelt's four laws often provide a gold standard on which models of binocular rivalry can be tested for biological realism and applied to discover general principles for visual perception, summarizing the core relationships between variations in stimulus characteristics and resultant rivalrous dynamics (Levelt, 1965). These propositions have evolved into a more modern form over the last half century, accumulating new insights over a wide array of modern experiments that are summarized in (Brascamp et al., 2015), and we address each in turn using our model network dynamics. While experimental studies involving Levelt's laws typically consider the impact of stimulus strength based on the relative contrast, density, or blur of contours in each monocular stimulus for relatively simple images, in our model we consider the stimulus strength to be directly determined by the mean feedforward input for a downstream neuron. While there is not a single notion of stimulus strength across all experiments, meaning alternative perturbations in the stimuli can predictably effect rivalrous dynamics, our choice allows us to precisely probe

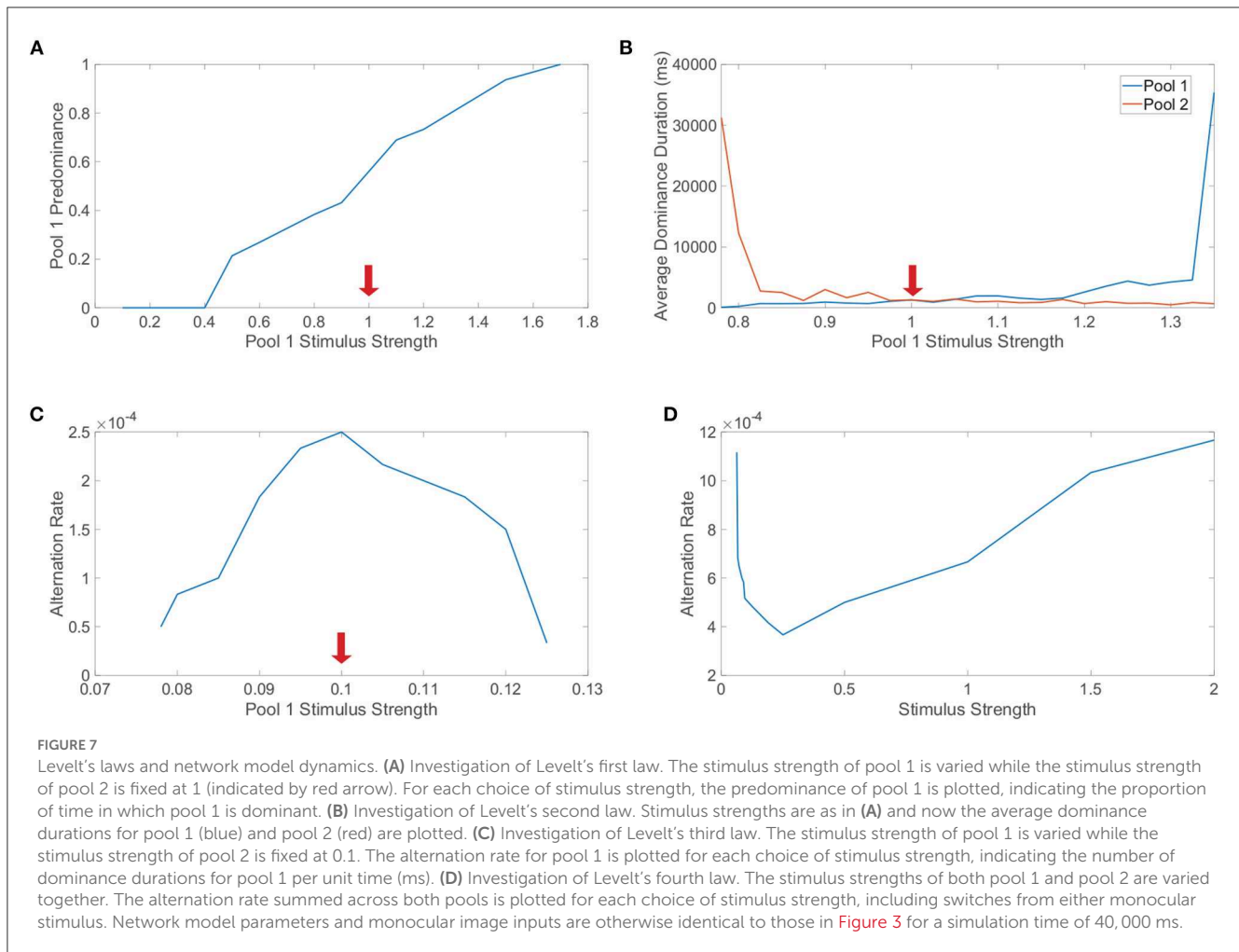
the strength of any stimulus in a unified manner, including complex natural scenes as well as simple gratings.

Levelt's first law posits that *increasing the strength of one monocular stimulus only will increase the predominance of that particular monocular stimulus*. Note that the predominance of a specific monocular stimulus is the proportion of viewing (simulation) time in which that monocular stimulus is dominant. In Figure 7A, we explore Levelt's first law by varying the monocular stimulus strength, m_0 , for the first pool only and fixing the monocular stimulus strength for the second pool. Across a sequence of such simulations, we plot the predominance of the first pool, which indeed increases with its monocular stimulus strength for sufficiently large feedforward drive; this trend holds over a range of m_0 values that broadly surrounds the fixed monocular stimulus strength for the second pool. Intuitively, Levelt's first law agrees with the idea that the pool with the stronger monocular stimulus should be dominant more often since it is predisposed to suppressing the competing pool.

Proceeding in turn, Levelt's second law postulates that *increasing the difference between the two monocular stimulus strengths will primarily increase the average dominance duration for the percept corresponding to the stronger monocular stimulus strength*. We investigate Levelt's second law in Figure 7B by again varying the monocular stimulus strength for only the first pool and now plotting the resultant average dominance durations for each pool across the corresponding simulations. We observe that when the second pool has a stronger monocular stimulus and the monocular stimulus strength for the first pool is decreased, the average dominance duration for the second pool markedly increases with minimal change in the average dominance duration for the first pool. Likewise, when the stimulus strength for the second pool is less than the stimulus strength for the first pool, as the stimulus strength for the first pool is increased, the average dominance duration for the first pool increases with only a minor change in the average dominance duration for the second pool. Together, these observations conform precisely to the prescription provided by Levelt's second law.

Levelt's third law proposes that *increasing the difference between the two monocular stimulus strengths will reduce the alternation rate*. Note that here the alternation rate refers to the total number of perceptual switches that occur per unit time. We explore Levelt's third law in Figure 7C by once again varying the monocular stimulus strength for the first pool while fixing that of the second pool, plotting the alternation rate in each case. As expected from Levelt's third proposition, the highest alternation rate occurs when the two monocular stimuli are of equal strength, with consistently decreasing alternation rate as the first pool stimulus strength is either further increased or decreased from the fixed second pool stimulus strength. One can argue that Levelt's third law is in fact equivalent to Levelt's second law, since when the average dominance durations for the two pools in sum are lowest, there will generally be the most alternations. However, the two laws are still often separated because they observe this shift in dynamics with a slightly different perspective and were more distinct in the context of the original non-updated version of Levelt's laws.

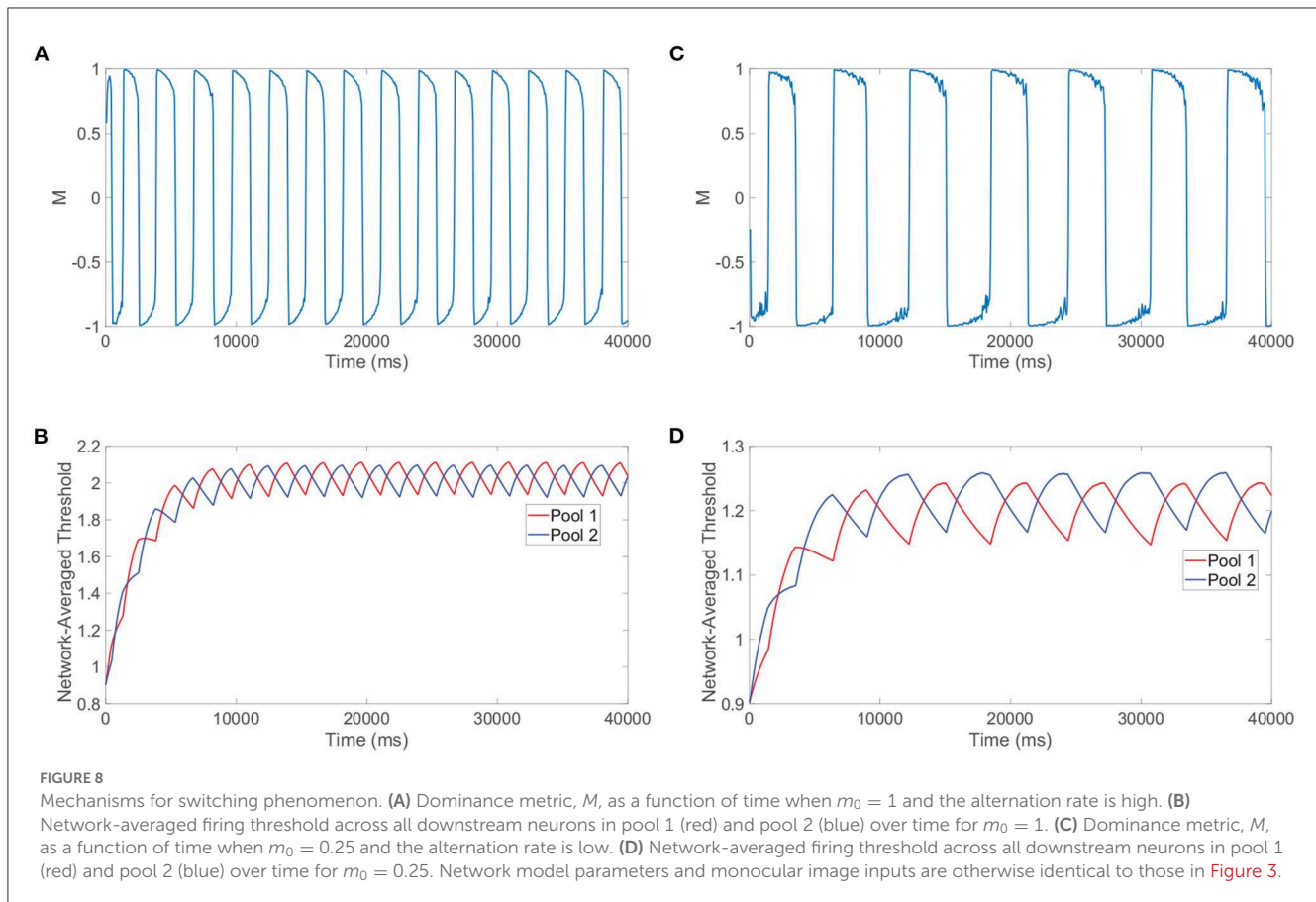
Finally, Levelt's fourth law asserts that *increasing the monocular stimulus strengths of each eye together, such that the two monocular*



stimulus strengths are equal in each case, will generally increase the alternation rate, except at particularly low monocular stimulus strengths for which this trend is reversed. To probe this final law, we vary together the monocular stimulus strengths for each pool, so the two downstream pools have the same average feedforward drive in a given trial, and plot the alternation rate for each choice of monocular stimulus strength in **Figure 7D**. We see that as the stimulus strength is increased beyond $m_0 = 0.25$, the alternation rate monotonically increases. On the other hand, for low stimulus strengths, we note that as the stimulus strength is instead decreased below 0.25, the alternation rate increases. These results fully agree with this modern form of Levelt's fourth law, including the more subtle shift in trend that occurs for low stimulus strengths. While the majority of studies demonstrate an increase in alternation rate with increasing stimulus strength, several model-driven explorations (Shapiro et al., 2009; Seely and Chow, 2011) as well as select experiments (Platonov and Goossens, 2013) alternatively exhibit an increase in alternation rate as stimulus strength is decreased over a range of weak strengths. It is likely that the regime of small stimulus strengths for which this reversal occurs is difficult to produce in experiments for specific notions of stimulus strength, causing most experimental settings to more naturally indicate that alternation rate increases with stimulus strength.

Beyond reproducing experimental observations, Levelt's fourth law and our related model analysis together support concrete mechanisms for the percept switching that depend on the strength of the stimuli. To demonstrate this, we compare the model network dynamics for two different stimulus strengths, with one in the classical Levelt's fourth law regime and the other in the reversal regime. We plot in **Figure 8A** the dominance metric dynamics for a relatively large stimulus strength of $m_0 = 1$ and depict the corresponding dynamics of the network-averaged thresholds for the two pools in **Figure 8B**. We do the same in **Figures 8C, D**, respectively, when the two downstream pools are each driven using a particularly small stimulus strength of $m_0 = 0.25$. As expected from Levelt's fourth law and **Figure 7D**, the dominance durations are generally longer when the stimulus strength is relatively small.

More informative to explaining the reversal at small stimulus strengths, however, is the difference in the saturation of the network-averaged thresholds in the two cases. For relatively high m_0 , it can be argued that the suppressed pool is responsible for the percept switch in what is often called the escape mechanism (Brascamp et al., 2015). In this case, decreasing the external drive strength results in longer periods of dominance (lower alternation rate) because generally the firing thresholds for neurons in the suppressed pool will need to drop yet lower to produce spikes and actively escape domination. As m_0 is decreased, the suppressed



pool has less total external drive, so the maximal voltage that a constituent neuron can achieve through fluctuations in its input current will be lower and hence the dynamic threshold will need to decrease further over the course of more time for a spike to be viable. In the regime of high m_0 , as exemplified by the steeply decreasing thresholds in Figure 8B, a necessary fixed decrease in firing threshold is achieved quite rapidly since the firing thresholds of the suppressed pool are relatively high initially and thus the first derivatives in threshold are generally large in magnitude in accordance with Equation (1b).

On the other hand, for small m_0 below 0.25, it can be argued that the dominant pool is instead responsible for the percept switch in what is often called the release mechanism. Here, decreasing the external drive strength results in shorter periods of dominance (higher alternation rate) because generally the firing thresholds for neurons in the dominant pool will need to rise less for the spiking of the dominant neurons to be extinguished and thereby release the suppressed pool from domination. Decreasing the external drive results in shorter dominance durations because neurons in the dominant pool become unable to spike for a lower firing threshold that will take less time to reach *via* spike-frequency adaptation. This alternative mechanism corresponds to the relatively shallow declines in firing threshold exhibited in Figure 8D, since for sufficiently small m_0 the amount of time necessary for the suppressed pool thresholds to recover (escape) becomes excessively long and what instead begins to happen first is the thresholds of neurons in the dominant pool become so high that spiking is no longer feasible. From the perspective of

Equation (1b), a necessary fixed decrease in threshold for neurons in the suppressed pool is achieved very slowly since the firing thresholds of constituent neurons are now relatively low and thus the first derivatives in threshold are typically small in magnitude; therefore, the mechanism changes and the dominant neurons' inability to fire largely drives the percept switching for sufficiently small m_0 . We note that the exact time of the switch for either mechanism is random due to the fluctuations in neuronal inputs and irregular dynamics in the balanced regime, and when the approach of the firing threshold toward a critical value necessary for switching is especially slow in the case of the long dominance durations for $m_0 = 0.25$, there is a larger degree of variability in the dominance durations. Such a shift from the escape to release mechanism observed empirically in our detailed neuronal model is indeed carefully documented and substantiated by prior theoretical studies (Curtu et al., 2008), where rigorous analytical techniques and dynamical systems theory were utilized in the context of a two-population firing rate model to characterize the change in mechanism for low stimulation strengths.

2.5. Rivalry dynamics and receptive field structure

While up to this point we have demonstrated the successful dynamic encoding of monocular stimulus percepts during binocular rivalry and the adherence of our model activity to

canonical experimental constraints, we now turn to show that these results are robust to more biologically faithful choices of feedforward connectivity beyond the uniformly random structure assumed in the prior sections. In particular, receptive field structure in feedforward connectivity is ubiquitous throughout much of the visual system in the sense that downstream neurons are most stimulated by a range of stimuli with analogous features (Welker, 1976; Graziano and Gross, 1993; Wilson, 2001). This includes neurons in the retina, LGN, and V1, though the precise prototypical features of the receptive fields vary by brain area. In the LGN, for instance, neurons generally possess center-surround receptive fields, aggregating the output of upstream ganglion cells that together sample a spatially localized region of visual space, with inputs from the central circle and surrounding annulus of the receptive field having opposite (excitatory or inhibitory) effect (Hubel and Wiesel, 1960; Wiesel, 1960). In this case, the overall size of the spatial receptive field determines the spatial frequencies in visual space that are encoded and the center-surround structure facilitates the encoding of image edge information.

Prior work in the context of a single stimulus driving both eyes has showed that, compared to the uniformly random feedforward connectivity initially assumed in our two-layer model, spatially localized feedforward connectivity inspired by receptive fields can more accurately encode the dominant low and moderate frequency components composing most natural stimuli (Barranca et al., 2016). We focus next on the spatially localized nature of these receptive fields for simplicity and generality, noting that the inclusion of center-surround antagonism does not significantly alter the compressed encoding of images in the downstream network dynamics and still adheres to the requirements of compressive sensing theory if sufficient randomness is incorporated (Barranca and Zhu, 2018). To include spatial structure, we assume that a pixel in n -vector \mathbf{p}_z is mapped to a distinct (x, y) location with integer coordinates on a $[1, \sqrt{n}] \times [1, \sqrt{n}]$ Cartesian grid corresponding to its row and column location in the pixel matrix that produced vectorization \mathbf{p}_z . Each row of the feedforward connectivity matrices is then mapped to a distinct random location (x_i, y_i) on this grid, around which the receptive field of the i^{th} downstream neuron is centered. Incorporating both spatial localization and randomness in the spatial receptive fields, we assume the probability, P , that the i^{th} downstream neuron samples a pixel with spatial coordinates (x_j, y_j) is given by

$$P = \rho e^{-[(x_i - x_j)^2 + (y_i - y_j)^2] / 2\sigma^2}, \quad (4)$$

where ρ is the sampling probability if $(x_i, y_i) = (x_j, y_j)$, when the receptive field is centered at the location of a given pixel, and σ scales the spatial size of the receptive field, which is known to vary throughout the visual system (Hubel, 1995; Sceniak et al., 1999). Therefore, each feedforward connection is prescribed by a Bernoulli random variable, determined independently of all other connections, with success probability given by Equation (4).

We utilize the (ρ, σ) parameter choice shown previously to optimally encode a single 100×100 pixel natural scene stimulus in a single-pool network model. The generated feedforward connectivity matrices are sparse with connection density near 0.001, where $\sigma = 2.2$ produces moderately-sized receptive fields and $\rho = 0.92$ allows each pixel to be sampled

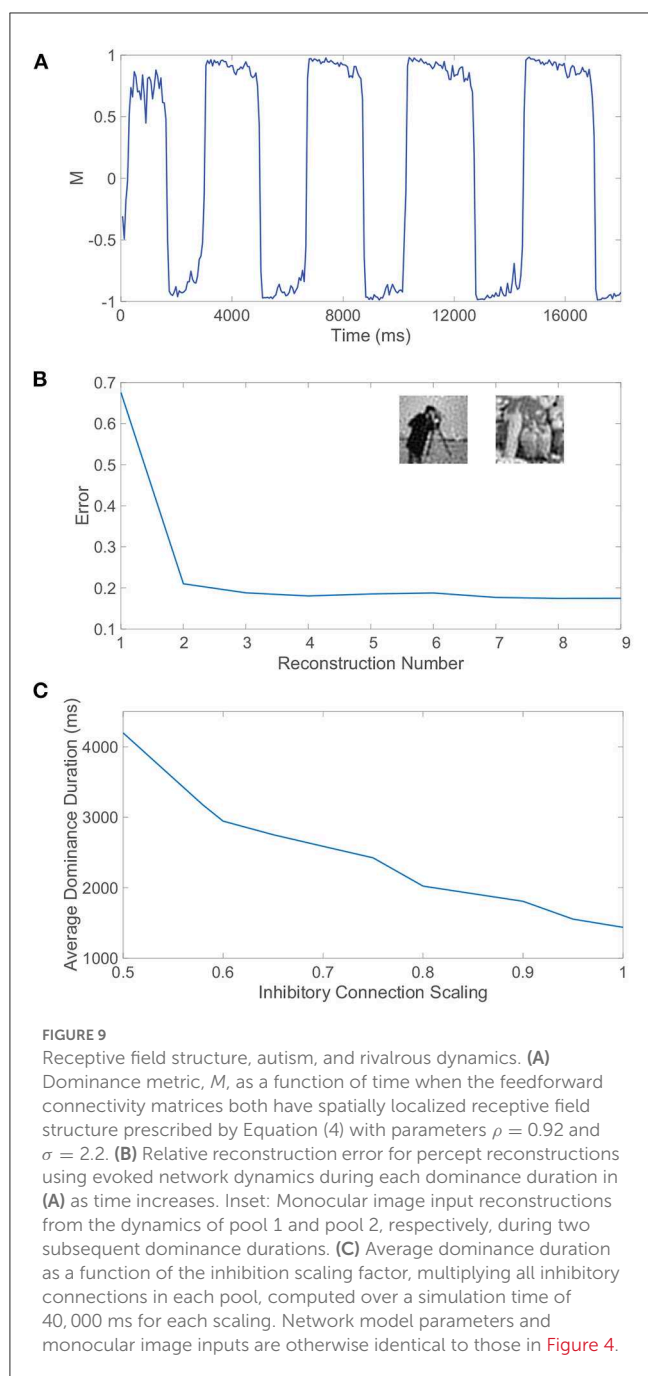
on average at least one time (Barranca et al., 2016). The resultant dominance metric dynamics are plotted in Figure 9A, which demonstrate clear rivalrous switching behavior and largely resemble the activity produced in the case of the uniformly random feedforward connectivity considered in the previous sections. What does change, however, is the quality of the monocular natural scene percept reconstructions during each period of dominance, as evidenced in Figure 9B. While there still exists an initial transient period in which neither stimulus is well-encoded, reconstructions using data from the subsequent dominance durations are significantly improved relative to those produced for the natural scenes considered in Figure 4B with uniformly random feedforward connectivity. The relative reconstruction errors are nearly halved, gravitating around a constant low value.

Sample reconstructions of the corresponding two monocular natural scene inputs are depicted in the inset of Figure 9B, visually demonstrating quite accurate encoding. Hence, the proposed mechanisms for binocular rivalry are robust to more realistic feedforward network architectures that are largely absent from prior model investigations. The continued presence of the initial transient period indicates that its entrenchment is beyond the feedforward connectivity structure, and the associated mixed percepts may in fact manifest as a result of initial information loss due to the sensory bottleneck or an intrinsic lag in processing via the non-linear network dynamics.

2.6. Relationship between autism and binocular rivalry

The excitation/inhibition imbalance hypothesis suggests that levels of excitation and inhibition are abnormal throughout many brain areas for individuals with autism (Rubenstein and Merzenich, 2003; Jamain et al., 2008), but clear markers for this imbalance in the autistic brain are still needed. At the same time, several experimental studies indicate that inhibition levels in the brain may be implicated in variations in both alternation rate and the prevalence of mixed percepts in binocular rivalry (Robertson et al., 2013; Dieter et al., 2017; Spiegel et al., 2019). Alternation rate may therefore serve as a non-invasive indicator of autism, and in light of evidence that the autistic brain exhibits reduced inhibition due to impaired GABA signaling (Robertson et al., 2016; Mentch et al., 2019), we explore the effect of reduced inhibition in our model on rivalrous dynamics.

To adjust the level of inhibition, we multiply all inhibitory connection strengths in our model, particularly parameters R_{EI} and R_{II} , by a scaling factor S_I . For each choice of inhibitory scaling factor, we record the dynamics of our model and compute the average dominance duration, plotting the result in Figure 9C. We observe that as the inhibition in the model decreases, the average dominance duration increases and the alternation rate correspondingly decreases. This monotonic trend strongly suggests that binocular rivalry is indeed weakened as a result of the reduced inhibition often associated with autism. In fact, if the inhibition strength is yet weaker than considered in Figure 9C, we begin to see trials in which either full winner-take-all dynamics manifest with no percept switching or interspersed long periods of time in which



the dominance metric is near 0 indicative of a mixed percept, which is consistent with the longer periods of mixed percepts reported for autistic individuals in experiments (Robertson et al., 2013).

It should be noted that the conditions for balanced dynamics in an isolated (or highly dominant) pool given by Equation (9) are still met under this scaling of inhibition, which has no impact on the quotient $\frac{|R_{EI}|}{|R_{II}|}$ in the inequality, and thus the diminished rivalry is theoretically not a direct result of balanced dynamics breaking down. It is potentially the case that the reduced inhibition results in weaker competition between the downstream pools and consequently the slow adaptation dynamics must evolve over a longer time scale to compensate. One possible explanation

according to Equation (8) is that the inhibitory neurons are expected to fire at a higher rate when their synaptic influence is weakened; so immediately following an alternation, the elevated firing thresholds of the inhibitory neurons in the newly suppressed pool will need to fall more to reach a fixed critical threshold for which a percept switch can occur and as a result more time will elapse until the next alternation.

3. Discussion

In the context of our newly formulated two-layer neuronal network model, we develop a framework for estimating the dynamic percepts in binocular rivalry based on explicit monocular stimulus reconstructions using measurements of network dynamics collected during each period of dominance. Our model suggests that to generate realistic rivalrous alternations it is sufficient to include irregular neuronal activity rooted in balanced network dynamics, competition between the downstream pools mediated by long-range connections from excitatory neurons in one pool to inhibitory neurons in the other, and slow fatigue *via* spike-frequency adaptation, showing that the removal of any one of these characteristics eliminates the dynamical features key to binocular rivalry. Moreover, beyond its novelty in the ability to directly incorporate natural scene monocular stimuli, our rivalry model is among the first to include complex network structure, such as feedforward connectivity and downstream compression, and uniquely produces distinct dynamics during the early transient period, which may correspond to the initial mixed percepts widely observed in experiments. We resolve the potential sensory bottleneck, where photoreceptor output is processed by the activity of smaller networks of neurons in downstream layers, by deriving an approximate linear input-output map embedded in the non-linear network dynamics and then utilizing compressive sensing theory to accurately recover a sequence of dominant monocular stimuli corresponding to the dynamic percepts. Incorporating additional facets of visual system structure observed empirically, such as spatial localization in receptive fields and adaptation present in excitatory neurons only, we underline the robustness of the proposed mechanisms for rivalrous dynamics and demonstrate strong agreement with canonical experimental findings that serve as litmus tests for rivalry models, such as Levelt's laws and gamma-distributed dominance durations.

While our study does address many features novel to theoretical investigations of binocular rivalry, experimental observations have unearthed additional factors that may play important roles in the phenomenon, such as top-down feedback from higher cortical areas and attention (Brown and Norcia, 1997; Zhang et al., 2011; Dieter et al., 2016; Li et al., 2017), and their dynamics may significantly modulate the lower-level processing primarily considered in this work. The primary visual cortex, as examined in our model, is the classical area implicated in generating perceptual bistability through what is termed eye-based rivalry, arising from competition between neurons driven by monocular inputs. While eye-based rivalry is well-supported experimentally (Xu et al., 2016), there is compelling evidence of stimulus-based rivalry, where the dynamics of binocular and feature-selective neurons in higher visual system areas are instead associated with perceptual changes

in binocular rivalry (Leopold and Logothetis, 1996). While we largely remain agnostic on this issue, it may be the case that the two proposed types of rivalry together yield a more complete explanation of the true phenomenon (Lee and Blake, 1999; Tong, 2001).

Beyond the bistability examined in this work, alternative perceptual behaviors closely related to binocular rivalry are possible and may be studied using analogous techniques. Multistable rivalry between more than two percepts can manifest when parts of the two monocular stimuli may be divided and recombined into various coherent images, generating perceptual alternations between the original monocular stimuli as well as the meaningfully regrouped coherent images (Kovacs et al., 1996; Suzuki and Grabowecy, 2002; Sterzer et al., 2009; Golubitsky et al., 2019). While the use of additional fine-grained spatial structure in the recurrent connectivity with additional pools may generalize our model to this setting, many other forms of visual multistability often involve multiple mutually exclusive three-dimensional interpretations of a single static two-dimensional image due to a lack of depth or orientation cues (Kornmeier and Bach, 2012), for which alternative model settings may be more appropriate in studying their mechanisms.

As multistable perception is documented across sensory systems (Holcombe and Seizova-Cajic, 2008; Hupe et al., 2008; Zhou and Chen, 2009), possibly manifesting from universal organizing principles, computational models have taken diverse creative approaches to granting important mechanistic insights. In the context of audition (Nguyen et al., 2020), for example, one recent study utilized a model featuring two noisy accumulators associated with different ambiguous stimuli, similar to how the dominance metric in this work incorporates firing rate data corresponding to the two monocular stimuli in order to produce the choice of visual percept. However, in the context of auditory bistable perception, the accumulators were forced by data-driven Poisson distributed inputs based on spike count data from the primary auditory cortex, showing that when the saturation strength, akin to adaptation in the present work, is close to the boundary that yields noise-driven attractor dynamics, perceptual alternations agreeing with key experimental observations manifest without the direct use of competition. Another recent study presented a compelling alternative mechanistic framework for binocular rivalry, replacing slow microscopic adaptation with an accumulation process carried out by pools of neurons in the first layer that fed into pools of competing neurons in the second layer (Cao et al., 2021). This alternative framework used out-of-equilibrium dynamics to reproduce common experimental findings and, beyond most existing mechanistic models, further agreed with detailed experimental observations in that it both demonstrated sequential dependence in dominance durations and preserved the statistics of dominance durations even when the contrast of the monocular stimuli was broadly adjusted.

In contrast to binocular rivalry, when the two monocular stimuli are instead sufficiently similar, binocular fusion may occur and result in a constant single image percept, with hysteretic transitions between rivalry and fusion as the similarity of the monocular stimuli varies (Nelson, 1975; Buckthought et al., 2008). How the competition mechanism is counteracted to shift from

rivalry to fusion in a unified manner is an important issue (Wilson, 2017) and may be further explored in the two-pool context in this work, for which it is possible to directly manipulate the monocular stimuli and measure the resulting percept given a reconstruction paradigm. While we use a particular monocular stimulus recovery method based on the dynamics of the rivalry metric and compressive sensing techniques applied to the derived input-output mapping given by Equation (2) for the dominant pool, it may also be possible to produce a reasonable percept based on the average of reconstructions generated using the input-output mappings for both the dominant and suppressed pools. Intuitively, for rivalrous dynamics it is likely that the reconstruction from the suppressed pool would only add minimal noise to the reconstruction generated from the dominant pool. In the case of fusion, this averaging approach could coherently generate a single fused percept and potentially yield a more unified understanding of these perceptual phenomena. In a similar vein, the impact of monocular contrasts in fusion for similar images and rivalry for highly distinct stimuli may be analyzed in this context (Wilson, 2017).

There is both theoretical and experimental evidence that the spreading of perceptual waves during alternations in dominance is associated with binocular rivalry (Wilson et al., 2001). In a perceptual transition, a perceptual wave pattern associated with a specific stimulus manifests locally and then spatially spreads with time to eventually eclipse the competing pattern, closely aligning with the propagation speed of traveling wave dynamics of cortical activity in V1 functional magnetic resonance imaging data (Lee et al., 2007). While feedforward connectivity with spatially localized structure akin to receptive fields was examined in Section 2.5, it would be informative to also provide a spatial arrangement in the recurrent connectivity for the two competing neuronal pools, potentially connecting neurons with analogous orientation selectivity, and examine when as well as how this imbues the system with spreading waves of dominance in perceptual alternations for different classes of monocular stimuli.

Giving further credence to empirical evidence that individuals with autism demonstrate weaker binocular rivalry (Robertson et al., 2013; Spiegel et al., 2019), we have systematically demonstrated that the percept alternation rate in binocular rivalry decreases as the network inhibitory connections diminish in strength akin to the reduced impact of GABA hypothesized to be present in the autistic brain (Robertson et al., 2016; Mentch et al., 2019). At the same time, there is still great debate regarding the excitation/inhibition imbalance hypothesis for autism (Happe et al., 2006; Dinstein et al., 2012), and further investigation, both theoretical and experimental, is warranted. Some experimental measurements suggest that the autistic brain is alternatively subject to extraneous excitation (Vattikuti and Chow, 2010) and genetic studies associating excitation or inhibition with autism generally fail to demonstrate a clear directional impact on synaptic transmission (Chao et al., 2010), indicating that continued interplay between experimental advances and mathematical theory is necessary to provide a more parsimonious characterization of the autistic brain.

Though we assumed the activity of the downstream neurons to be prescribed by integrate-and-fire model dynamics, we

could similarly investigate whether our framework for rivalrous alternations is robust to more biologically realistic single-neuron models, such as the exponential integrate-and-fire and Hodgkin-Huxley models (Hodgkin and Huxley, 1952; Barranca et al., 2014a). While such models present significant conceptual obstacles in deriving new network input-output mappings crucial to percept reconstructions and may not be analytically tractable, we have developed a methodology for constructing data-driven input-output maps across two layers of excitatory neurons (Barranca et al., 2021) and preliminary work indicates this approach may be extended to balanced networks. Considering that binocular rivalry has provided a unique and fruitful lens through which neuronal computation in the visual system can be understood, linking a dynamic percept to constant external drive, we expect that going forward analogous model-based analysis may help to characterize mechanisms for stimulus encoding and multistability in other sensory systems (Holcombe and Seizova-Cajic, 2008; Hupe et al., 2008; Zhou and Chen, 2009).

4. Methods

4.1. Network input-output mapping derivation

Under the assumption of irregular neuronal network activity prescribed by Equation (1), in this section we derive a linear input-output relationship crucial to our framework for reconstructing monocular stimulus inputs in Section 2.2. To uncover the underlying network mapping, we estimate the expected voltage for each downstream neuron in the long-time limit. For analytical tractability, we begin by approximating the net effect of all recurrent and feedforward inputs into a given downstream neuron. In considering the recurrent interactions, it is useful to observe that the input into a particular neuron is a spike train summed over the action potentials generated from a relatively large number of neighboring neurons. Since the firing events of neurons in the balanced regime are only weakly correlated, the summed spike train input over a large number of incoming neuronal impulses asymptotically approaches a Poisson point process (Cinlar, 1972). Moreover, because the external input into each pool is constant, we arrive at a simpler statistical characterization of the neuronal inputs. Note that while we derive this input-output mapping for the downstream neurons in the first pool for concreteness, the corresponding mapping for neurons in the second pool is given by interchanging the 1 and 2 subscripts in the equations below.

Since the voltage of each downstream neuron is reset to V^{Re} upon firing, we consider Equation (1) with initial condition $v_{1k}^i(t = 0) = V^{Re}$ for $k = E, I$. Assuming Poisson spike train inputs as described above, the solution to the corresponding initial value problem gives the subthreshold membrane potential trajectory for the i^{th} neuron of the k^{th} population in the first pool and may be expressed as

$$v_{1k}^i(t) = V^{Re} + \sum_{\substack{j=1 \\ j \neq i}}^{N_E} R_{1kE}^{ij} \Psi_{1E}^j(t) + \sum_{\substack{j=1 \\ j \neq i}}^{N_I} R_{1kI}^{ij} \Psi_{1I}^j(t)$$

$$+ \sum_{\substack{j=1 \\ j \neq i}}^{N_E} C_{2kE}^{ij} \Psi_{2E}^j(t) + \left(\frac{1 - e^{-gLt}}{gL} \right) \sum_j F_{1k}^{ij} p_1^j \quad (5a)$$

$$\Psi_{z_l}^j(t) = \sum_{s=1}^{T_{z_l}^j(t)} e^{-gL(t - U_{z_l s}^j(t))}, \quad (5b)$$

where $R_{1kl}^{ij} \cdot \Psi_{1l}^j(t)$ yields the total recurrent spike train input from the j^{th} neuron of the l^{th} population in the first pool into the i^{th} neuron of the k^{th} population in the first pool at time t and $T_{z_l}^j(t)$ denotes the total number of spikes transmitted by the j^{th} neuron of the l^{th} population in the z^{th} pool through time t . Note that here $T_{z_l}^j(t)$ is described approximately by a Poisson distribution with average number of events $m_{z_l}^j t$, where $m_{z_l}^j$ is the long-time average firing rate of the j^{th} neuron of the l^{th} population in the z^{th} pool.

Corresponding to $T_{z_l}^j(t)$, the spike train input from the j^{th} neuron of the l^{th} population in the z^{th} pool has spike times denoted by $U_{z_l s}^j(t)$, for $s = 1, 2, \dots$, which we assume are uniformly distributed in the time interval $[0, t]$ based on the irregularity of firing events. Hence, the term in Equation (5) given by $Y_{z_l s}^j(t) = e^{-gL(t - U_{z_l s}^j(t))}$ is a random variable that takes on values in the interval $[e^{-gLt}, 1]$. Since the probability density function for random variable $U_{z_l s}^j(t)$ is $P_{U_{z_l s}^j(t)}(u) = 1/t$ in its support, it follows that random variable $Y_{z_l s}^j(t)$ has probability density function $P_{Y_{z_l s}^j(t)}(y) = 1/(gLty)$ for $y \in [e^{-gLt}, 1]$ and expected value $(1 - e^{-gLt})/(gLt)$.

Since $\Psi_{z_l}^j(t)$ is a sum of nearly independent identically distributed random variables, it has expected value $\frac{1 - e^{-gLt}}{gLt} m_{z_l}^j t$ for time horizon t . Temporarily disregarding the impact of the reset condition, the expected voltage of the i^{th} neuron of the k^{th} population in the first pool at time t is thus approximately

$$\tilde{v}_{1k}^i(t) = V^{Re} + \frac{1 - e^{-gLt}}{gL} \cdot \left(\sum_j F_{1k}^{ij} p_1^j + \sum_{\substack{j=1 \\ j \neq i}}^{N_E} R_{1kE}^{ij} m_{1E}^j + \sum_{\substack{j=1 \\ j \neq i}}^{N_I} R_{1kI}^{ij} m_{1I}^j + \sum_{\substack{j=1 \\ j \neq i}}^{N_E} C_{2kE}^{ij} m_{2E}^j \right).$$

As $t \rightarrow \infty$, we have that $\tilde{v}_{1k}^i(t)$ approaches

$$\tilde{v}_{1k}^i = V^{Re} + \frac{1}{gL} \cdot \left(\sum_j F_{1k}^{ij} p_1^j + \sum_{\substack{j=1 \\ j \neq i}}^{N_E} R_{1kE}^{ij} m_{1E}^j + \sum_{\substack{j=1 \\ j \neq i}}^{N_I} R_{1kI}^{ij} m_{1I}^j + \sum_{\substack{j=1 \\ j \neq i}}^{N_E} C_{2kE}^{ij} m_{2E}^j \right).$$

The true expected voltage, \tilde{v}_{1k}^i , will be lower than \tilde{v}_{1k}^i since upon reaching the dynamic threshold θ_{1k}^i , the voltage is instantaneously reset down to V^{Re} . As a result of this non-linearity, since the change in voltage due to a given action potential is $-(\theta_{1k}^i - V^{Re})$ occurring with rate m_{1k}^i , it follows that the true long-time expected voltage for the i^{th} neuron in the k^{th} population of the first pool, \tilde{v}_{1k}^i , is approximately

$$\tilde{v}_{1k}^i = V^{Re} + \frac{1}{gL} \cdot \left(\sum_j F_{1k}^{ij} p_1^j \right)$$

$$\begin{aligned}
 & + \sum_{\substack{j=1 \\ j \neq i}}^{N_E} R_{1_{kE}}^{ij} m_{1E}^j + \sum_{\substack{j=1 \\ j \neq i}}^{N_I} R_{1_{kI}}^{ij} m_{1I}^j \\
 & + \sum_{\substack{j=1 \\ j \neq i}}^{N_E} C_{2_{kE}}^{ij} m_{2E}^j - m_{1k}^i (\bar{\theta}_{1k}^i - V^{Re}), \tag{6}
 \end{aligned}$$

where $\bar{\theta}_{1k}^i$ is the time-averaged firing threshold for the i^{th} neuron in the k^{th} population of the first pool. Linking the evoked downstream network dynamics to the monocular image input given by \mathbf{P}_1 , Equation (6) is the linear input-output mapping that we will leverage to estimate the encoded percepts in Section 2.2 and is expressed in matrix form in Equation (2).

With this mapping in hand, we next derive conditions on the model parameters that theoretically produce balanced dynamics in the dominant pool. The excitatory neurons in the suppressed pool are sufficiently quiescent such that the cross-pool inputs in Equation (6) determined by matrix \mathbf{C}_2 have relatively little impact on the expected voltage offset from V^{Re} , and thus we derive balance conditions for the dominant pool ignoring the competitive long-range inputs included in our full model. Viewing the dynamics at the population level, since each neuron in the l^{th} population of the first pool is expected to fire at rate given by the long-time population-averaged firing rate m_{1l} and each neuron in the k^{th} population of the first pool is expected to receive K incoming recurrent connections from the l^{th} population of the first pool with individual connection strength R_{kl}/\sqrt{K} , the expected total recurrent input from the l^{th} population in the first pool into a neuron in the k^{th} population of the first pool is approximately $R_{kl}m_{1l}\sqrt{K}$. Similarly, we assume the expected feedforward input into a downstream neuron in the k^{th} population of an arbitrary pool is of magnitude $\sqrt{K}f_k m_0$, with $\mathcal{O}(1)$ parameters m_0 and f_k respectively scaling the overall and relative feedforward drive strengths for the excitatory and inhibitory populations. As a result, taking the expectation of Equation (6) over all network realizations, we approximate the expected offset in voltage from V^{Re} due to both recurrent and feedforward inputs for a downstream neuron in the k^{th} population of the first pool as

$$d_{1k} = \sqrt{K} \frac{1}{g_L} (f_k m_0 + R_{kE} m_{1E} + R_{kI} m_{1I}). \tag{7}$$

Seeking irregular activity indicative of balance within the dominant pool, the expected voltage for a given downstream neuron must remain finite as $K \rightarrow \infty$ in the large-network limit. Based on Equation (7), which neglects the weak cross-pool drive and assumes statistically equivalent constant feedforward drive in the two pools for fairness in rivalry, it must be that $f_k m_0 + R_{kE} m_{1E} + R_{kI} m_{1I}$ is $\mathcal{O}(1/\sqrt{K})$ and vanishes as $K \rightarrow \infty$ (van Vreeswijk and Sompolinsky, 1998). The resultant constraints for both the excitatory and inhibitory populations composing the first pool provide a system of two linear equations with solution

$$m_{1E} = \frac{|R_{II}|f_E - |R_{EI}|f_I}{R_{IE}|R_{EI}| - R_{EE}|R_{II}|} m_0 \tag{8a}$$

$$m_{1I} = \frac{R_{IE}f_E - R_{EE}f_I}{R_{IE}|R_{EI}| - R_{EE}|R_{II}|} m_0, \tag{8b}$$

indicating a linear scaling of the population-averaged firing rates with the overall feedforward input strength parameter m_0 .

Requiring that both the excitatory and inhibitory population-averaged firing rates are non-negative and finite in the balanced dynamical regime, Equation (8) provides parameter bounds referred to as balance conditions,

$$\frac{f_E}{f_I} > \frac{|R_{EI}|}{|R_{II}|} > \frac{R_{EE}}{R_{IE}}, \tag{9}$$

which are theoretically necessary for balanced dynamics in the dominant pool in the large-network limit. The linear scaling of downstream neuronal firing rates with feedforward input strength demonstrated by Equation (8) holds theoretically on the population level when the balance conditions are satisfied and in Figure 1E we empirically examine the population-averaged firing rates for the model network in response to increasingly large feedforward inputs by increasing the overall feedforward scaling strength m_0 across a family of simulations. We observe that as the mean feedforward drive is increased, the average firing rate of both the excitatory and inhibitory populations in a given pool linearly increases over a wide range of scalings, agreeing with the estimate computed in the large-network limit and corroborating the accuracy of the derived input-output mapping.

While the original balanced network theory was derived in the context of constant and homogeneous excitatory external inputs driving the two populations in a single pool of binary neurons (van Vreeswijk and Sompolinsky, 1996), it is important to emphasize that in our two-layer I&F network model the excitatory external input vectors are instead heterogeneous and determined by the pixels composing the monocular image stimuli. Nonetheless, as long as the stimulus strength scalings for the excitatory and inhibitory populations in a given pool are such that $f_E > f_I$ for these heterogeneous external inputs, the classical balance conditions given by Equation (9) hold on average and balanced dynamics are primarily preserved in the dominant pool.

4.2. Compressive sensing theory

Compressive sensing theory shows that for sparse data, the amount of measurements necessary for a successful reconstruction is determined by the number of dominant non-zero components in the data (Candes et al., 2006; Donoho, 2006). This suggests that optimally reconstructing sparse data from relatively few samples requires selecting the sparsest viable reconstruction that agrees with the collected information, since the resultant signal is most compressible. By leveraging the sparse structure of signals, CS offers a significant improvement in efficiency from the classical Shannon-Nyquist theorem, which asserts that the sampling rate should be determined by the full bandwidth of the data (Shannon, 1949). Considering that common signals and sensory stimuli, such as scenes, soundwaves, and odorants, are each sparse in a certain domain (Field, 1994; Markram et al., 1997), CS theory has amassed numerous and broad scientific applications (Lustig et al., 2007; Dai et al., 2009; Berger et al., 2010; Gross et al., 2010).

The reconstruction of time-invariant data from a small number of samples generates a highly underdetermined linear system. For

an n -component signal, \mathbf{x} , a total of m discrete samples of \mathbf{x} can be represented by \mathbf{Ax} , where \mathbf{A} is an $m \times n$ measurement matrix composed of rows that each prescribe a set of measurement weights. This yields an m -component measured signal, \mathbf{b} , where $m \ll n$. Assuming that signal \mathbf{x} is sufficiently sparse, CS theory shows for a large class of measurement matrices that minimizing $\|\mathbf{x}\|_{\ell_1} = \sum_{i=1}^n |x_i|$ yields the sparsest reconstruction consistent with the measurements (Candes and Wakin, 2008; Bruckstein et al., 2009). The resultant optimization problem

$$\arg \min_{\mathbf{x} \in \mathbb{R}^n} \|\mathbf{x}\|_{\ell_1} \text{ subject to } \mathbf{Ax} = \mathbf{b} \quad (10)$$

is well-researched and can be efficiently solved using numerous fast numerical methods, thus yielding a tractable reconstruction methodology for high-dimensional data with a sparse representation (Tropp and Gilbert, 2007; Donoho and Tsaig, 2008).

The recovered signal exhibits variations depending on the optimization algorithm used, but the minimal relative reconstruction errors generated across methods are quite comparable for our work. Moreover, if signal \mathbf{x} is not sparse in the sampled domain and is instead sparse under a transform, L , then the linear system $\phi \hat{\mathbf{x}} = \mathbf{b}$, where $\phi = \mathbf{AL}^{-1}$ and $\hat{\mathbf{x}} = L\mathbf{x}$, can be considered similarly. In the case of the natural scenes analyzed in our model, the monocular image input is indeed non-sparse in the original pixel domain but is instead sparse in a frequency-based domain. For concreteness, we utilize the two-dimensional discrete-cosine transform to take advantage of the sparse stimulus structure underlying our percept reconstructions (Heil and Walnut, 1989; Barranca et al., 2014b).

In determining an appropriate sampling scheme, it is important to note that measurement matrices exhibiting little correlation among their columns and well-preserving signal magnitudes, such as those containing independent identically distributed elements, are typically feasible candidates (Baraniuk, 2007; Candes and Wakin, 2008). While diverse matrices demonstrating sufficient randomness in their structure have been proven to suitably exhibit these properties (Candes et al., 2006; Candes and Wakin, 2008), there are other sampling matrices with more complex structure that are highly successful in CS signal reconstructions but not analytically tractable. Improved sampling matrices may

be generated, for example, from experimental insights or machine learning techniques (Barranca, 2021). As a result, numerous sampling schemes are amenable to CS recovery and they may be adapted based on logistical constraints or a priori knowledge of signal characteristics.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

Funding

This work was supported by a Swarthmore Faculty Research Support Grant and Eugene M. Lang Faculty Fellowship (VB).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Abbott, L. (1999). Lapicque's introduction of the integrate-and-fire model neuron (1907). *Brain Res. Bull.* 50, 303–304.
- Alais, D., and Melcher, D. (2007). Strength and coherence of binocular rivalry depends on shared stimulus complexity. *Vision Res.* 47, 269–279. doi: 10.1016/j.visres.2006.09.003
- Anderson, J. R., Jones, B. W., Watt, C. B., Shaw, M. V., Yang, J.-H., Demill, D., et al. (2011). Exploring the retinal connectome. *Mol. Vis.* 17, 355–379. Available online at: <http://www.molvis.org/molvis/v17/a41>
- Augustin, M., Ladenbauer, J., and Obermayer, K. (2013). How adaptation shapes spike rate oscillations in recurrent neuronal networks. *Front. Comput. Neurosci.* 7, 9. doi: 10.3389/fncom.2013.00009
- Baker, D. H., and Graf, E. W. (2009). Natural images dominate in binocular rivalry. *Proc. Natl. Acad. Sci. U.S.A.* 106, 5436–5441. doi: 10.1073/pnas.0812860106
- Baraniuk, R. (2007). Compressive sensing. *IEEE Signal Process. Mag.* 128–133. doi: 10.1109/MSP.2007.4286571
- Barlow, H. B. (1961). "The coding of sensory messages," in *Current Problems in Animal Behaviour*, eds W. H. Thorpe and O. L. Zangwill (Cambridge University Press), 331–360.
- Barlow, H. B. (1981). The ferrier lecture, 1980. critical limiting factors in the design of the eye and visual cortex. *Proc. R. Soc. Lond. B Biol. Sci.* 212, 1–34.
- Barranca, V. J. (2021). Neural network learning of improved compressive sensing sampling and receptive field structure. *Neurocomputing* 455, 368–378. doi: 10.1016/j.neucom.2021.05.061
- Barranca, V. J. (2022). Reconstruction of sparse recurrent connectivity and inputs from the nonlinear dynamics of neuronal networks. *J. Comput. Neurosci.* 51, 43–58. doi: 10.1007/s10827-022-00831-x

- Barranca, V. J., Hu, Y., Porterfield, Z., Rothstein, S., and Xuan, A. (2021). Data-driven reconstruction and encoding of sparse stimuli across convergent sensory layers from downstream neuronal network dynamics. *SIAM J. Appl. Dyn. Syst.* 20, 2602–2629. doi: 10.1137/21M1403114
- Barranca, V. J., Huang, H., and Li, S. (2019). The impact of spike-frequency adaptation on balanced network dynamics. *Cogn. Neurodyn.* 13, 105–120. doi: 10.1007/s11571-018-9504-2
- Barranca, V. J., Johnson, D. C., Moyher, J. L., Sauppe, J. P., Shkarayev, M. S., Kovačić, G., et al. (2014a). Dynamics of the exponential integrate-and-fire model with slow currents and adaptation. *J. Comput. Neurosci.* 37, 161–180. doi: 10.1007/s10827-013-0494-0
- Barranca, V. J., Kovačić, G., Zhou, D., and Cai, D. (2014b). Sparsity and compressed coding in sensory systems. *PLoS Comput. Biol.* 10, e1003793. doi: 10.1371/journal.pcbi.1003793
- Barranca, V. J., Kovačić, G., Zhou, D., and Cai, D. (2016). Improved compressive sensing of natural scenes using localized random sampling. *Sci. Rep.* 6, 31976. doi: 10.1038/srep31976
- Barranca, V. J., and Zhou, D. (2019). Compressive sensing inference of neuronal network connectivity in balanced neuronal dynamics. *Front. Neurosci.* 13, 1101. doi: 10.3389/fnins.2019.01101
- Barranca, V. J., and Zhu, X. G. (2018). A computational study of the role of spatial receptive field structure in processing natural and non-natural scenes. *J. Theor. Biol.* 454, 268–277. doi: 10.1016/j.jtbi.2018.06.011
- Benda, J., and Herz, A. V. (2003). A universal model for spike-frequency adaptation. *Neural Comput.* 15, 2523–2564. doi: 10.1162/089976603322385063
- Berger, C. R., Wang, Z., Huang, J., and Zhou, S. (2010). Application of compressive sensing to sparse channel estimation. *Comm. Mag.* 48, 164–174. doi: 10.1109/MCOM.2010.5621984
- Binas, J., Rutishauser, U., Indiveri, G., and Pfeiffer, M. (2014). Learning and stabilization of winner-take-all dynamics through interacting excitatory and inhibitory plasticity. *Front. Comput. Neurosci.* 8, 68. doi: 10.3389/fncom.2014.00068
- Blake, R., and Fox, R. (1974). Adaptation to invisible gratings and the site of binocular rivalry suppression. *Nature* 249, 488–490.
- Blake, R., and Logothetis, N. K. (2002). Visual competition. *Nat. Rev. Neurosci.* 3, 13–21. doi: 10.1038/nrn701
- Blake, R., Yang, Y., and Westendorf, D. (1991). Discriminating binocular fusion from false fusion. *Invest. Ophthalmol. Vis. Sci.* 32, 2821–2825.
- Blasdel, G. (1992). Differential imaging of ocular dominance and orientation selectivity in monkey striate cortex. *J. Neurosci.* 12, 3115–3138.
- Bobin, J., Starck, J.-L., and Ottensamer, R. (2008). Compressed sensing in astronomy. *J. Select. Topics Signal Process.* 2, 718–726. doi: 10.1109/JSTSP.2008.2005337
- Bock, E. A., Fesi, J. D., Baillet, S., and Mendola, J. D. (2019). Tagged meg measures binocular rivalry in a cortical network that predicts alternation rate. *PLoS ONE* 14, e0218529. doi: 10.1371/journal.pone.0218529
- Boerlin, M., Machens, C. K., and Deneve, S. (2013). Predictive coding of dynamical variables in balanced spiking networks. *PLoS Comput. Biol.* 9, e1003258. doi: 10.1371/journal.pcbi.1003258
- Borofsky, T., Barranca, V. J., Zhou, R., von Trentini, D., Broadrup, R. L., and Mayack, C. (2020). Hive minded: like neurons, honey bees collectively integrate negative feedback to regulate decisions. *Anim. Behav.* 168, 33–44. doi: 10.1016/j.anbehav.2020.07.023
- Bosten, J. M., Goodbourn, P. T., Lawrance-Owen, A. J., Bargary, G., Hogg, R. E., and Mollon, J. D. (2015). A population study of binocular function. *Vision Res.* 110(Pt A), 34–50. doi: 10.1016/j.visres.2015.02.017
- Brascamp, J. W., Klink, P. C., and Levelt, W. J. (2015). The 'laws' of binocular rivalry: 50 years of Levelt's propositions. *Vision Res.* 109(Pt A), 20–37. doi: 10.1016/j.visres.2015.02.019
- Brecht, M., and Sakmann, B. (2002). Dynamic representation of whisker deflection by synaptic potentials in spiny stellate and pyramidal cells in the barrels and septa of layer 4 rat somatosensory cortex. *J. Physiol.* 543(Pt 1), 49–70. doi: 10.1113/jphysiol.2002.018465
- Breese, B. (1909). Binocular rivalry. *Psychol. Rev.* 16, 410.
- Brette, R., Rudolph, M., Carnevale, T., Hines, M., Beeman, D., Bower, J. M., et al. (2007). Simulation of networks of spiking neurons: a review of tools and strategies. *J. Comput. Neurosci.* 23, 349–398. doi: 10.1007/s10827-007-0038-6
- Britten, K. H., Shadlen, M. N., Newsome, W. T., and Movshon, J. A. (1993). Responses of neurons in macaque MT to stochastic motion signals. *Vis. Neurosci.* 10, 1157–1169.
- Brown, D. A., and Adams, P. R. (1980). Muscarinic suppression of a novel voltage-sensitive K⁺ current in a vertebrate neurone. *Nature* 283, 673–676.
- Brown, R. J., and Norcia, A. M. (1997). A method for investigating binocular rivalry in real-time with the steady-state VEP. *Vision Res.* 37, 2401–2408.
- Bruckstein, A. M., Donoho, D. L., and Elad, M. (2009). From sparse solutions of systems of equations to sparse modeling of signals and images. *SIAM Rev.* 51, 34–81. doi: 10.1137/060657704
- Buckthrought, A., Kim, J., and Wilson, H. R. (2008). Hysteresis effects in stereopsis and binocular rivalry. *Vision Res.* 48, 819–830. doi: 10.1016/j.visres.2007.12.013
- Burkitt, A. N. (2006). A review of the integrate-and-fire neuron model: I. homogeneous synaptic input. *Biol. Cybern.* 95, 1–19. doi: 10.1007/s00422-006-0068-6
- Candes, E. J., Romberg, J. K., and Tao, T. (2006). Stable signal recovery from incomplete and inaccurate measurements. *Commun. Pur. Appl. Math.* 59, 1207–1223. doi: 10.1002/cpa.20124
- Candes, E. J., and Wakin, M. B. (2008). An introduction to compressive sampling. *Signal Process. Mag. IEEE* 25, 21–30. doi: 10.1109/MSP.2007.914731
- Cao, R., Pastukhov, A., Aleshin, S., Mattia, M., and Braun, J. (2021). Binocular rivalry reveals an out-of-equilibrium neural dynamics suited for decision-making. *eLife* 10, e61581. doi: 10.7554/eLife.61581.sa2
- Carter, O., van Swinderen, B., Leopold, D. A., Collin, S. P., and Maier, A. (2020). Perceptual rivalry across animal species. *J. Comp. Neurol.* 528, 3123–3133. doi: 10.1002/cne.24939
- Chao, H. T., Chen, H., Samaco, R. C., Xue, M., Chahrouh, M., Yoo, J., et al. (2010). Dysfunction in GABA signalling mediates autism-like stereotypies and Rett syndrome phenotypes. *Nature* 468, 263–269. doi: 10.1038/nature09582
- Cinlar, E. (1972). "Superposition of point processes," in *Stochastic Point Processes: Statistical Analysis, Theory, and Applications*, ed P. Lewis (New York, NY: Wiley), 549–606.
- Cohen, B. P., Chow, C. C., and Vattikuti, S. (2019). Dynamical modeling of multi-scale variability in neuronal competition. *Commun. Biol.* 2, 319. doi: 10.1038/s42003-019-0555-7
- Curtu, R., Shpiro, A., Rubin, N., and Rinzel, J. (2008). Mechanisms for frequency control in neuronal competition models. *SIAM J. Appl. Dyn. Syst.* 7, 609–649. doi: 10.1137/070705842
- Dai, W., Sheikh, M. A., Milenkovic, O., and Baraniuk, R. G. (2009). Compressive sensing DNA microarrays. *J. Bioinform. Syst. Biol.* 2009, 162824. doi: 10.1155/2009/162824
- Dayan, P. (1998). A hierarchical model of binocular rivalry. *Neural Comput.* 10, 1119–1135.
- Dieter, K. C., Brascamp, J., Tadin, D., and Blake, R. (2016). Does visual attention drive the dynamics of bistable perception? *Atten. Percept. Psychophys.* 78, 1861–1873. doi: 10.3758/s13414-016-1143-2
- Dieter, K. C., Sy, J. L., and Blake, R. (2017). Individual differences in sensory eye dominance reflected in the dynamics of binocular rivalry. *Vision Res.* 141, 40–50. doi: 10.1016/j.visres.2016.09.014
- Dinstein, I., Heeger, D. J., Lorenzi, L., Minshew, N. J., Malach, R., and Behrmann, M. (2012). Unreliable evoked responses in autism. *Neuron* 75, 981–991. doi: 10.1016/j.neuron.2012.07.026
- Donoho, D. L. (2006). Compressed sensing. *IEEE Trans. Inform. Theory* 52, 1289–1306. doi: 10.1109/TIT.2006.871582
- Donoho, D. L., and Tsai, Y. (2008). Fast solution of l_1 -norm minimization problems when the solution may be sparse. *IEEE Trans. Inform. Theory* 54, 4789–4812. doi: 10.1109/TIT.2008.929958
- Douglas, R. J., and Martin, K. A. (2004). Neuronal circuits of the neocortex. *Annu. Rev. Neurosci.* 27, 419–451. doi: 10.1146/annurev.neuro.27.070203.144152
- Dutour, É. (1760). Discussion d'une question d'optique. *l'académie des sciences. Mémoires de Mathématique et de physique presentés par Divers Savants* 3, 514–530.
- Emad, A., and Milenkovic, O. (2014). CaSPIAN: a causal compressive sensing algorithm for discovering directed interactions in gene networks. *PLoS ONE* 9, e90781. doi: 10.1371/journal.pone.0090781
- Ermentrout, B., Pascal, M., and Gutkin, B. (2001). The effects of spike frequency adaptation and negative feedback on the synchronization of neural oscillators. *Neural Comput.* 13, 1285–1310. doi: 10.1162/08997660152002861
- Ferster, D., Chung, S., and Wheat, H. (1996). Orientation selectivity of thalamic input to simple cells of cat visual cortex. *Nature* 380, 249–252.
- Field, D. J. (1994). What is the goal of sensory coding? *Neural Comput.* 6, 559–601.
- Field, G. D., and Chichilnisky, E. J. (2007). Information processing in the primate retina: circuitry and coding. *Annu. Rev. Neurosci.* 30, 1–30. doi: 10.1146/annurev.neuro.30.051606.094252
- Fiete, I. R., Fee, M. S., and Seung, H. S. (2007). Model of birdsong learning based on gradient estimation by dynamic perturbation of neural conductances. *J. Neurophysiol.* 98, 2038–2057. doi: 10.1152/jn.01311.2006
- Freeman, A. W. (2005). Multistage model for binocular rivalry. *J. Neurophysiol.* 94, 4412–4420. doi: 10.1152/jn.00557.2005

- Gallagher, R. M., and Arnold, D. H. (2014). Interpreting the temporal dynamics of perceptual rivalries. *Perception* 43, 1239–1248. doi: 10.1068/p7648
- Ganmor, E., Segev, R., and Schneidman, E. (2011). The architecture of functional interaction networks in the retina. *J. Neurosci.* 31, 3044–3054. doi: 10.1523/JNEUROSCI.3682-10.2011
- Gao, R., and Penzes, P. (2015). Common mechanisms of excitatory and inhibitory imbalance in schizophrenia and autism spectrum disorders. *Curr. Mol. Med.* 15, 146–167. doi: 10.2174/1566524015666150303003028
- Gershman, S. J., Vul, E., and Tenenbaum, J. B. (2012). Multistability and perceptual inference. *Neural Comput.* 24, 1–24. doi: 10.1162/NECO_a_00226
- Gilbert, C. D. (1992). Horizontal integration and cortical dynamics. *Neuron* 9, 1–13.
- Golubitsky, M., Zhao, Y., Wang, Y., and Lu, Z. L. (2019). Symmetry of generalized rivalry network models determines patterns of interocular grouping in four-location binocular rivalry. *J. Neurophysiol.* 122, 1989–1999. doi: 10.1152/jn.00438.2019
- Graziano, M. S., and Gross, C. G. (1993). A bimodal map of space: somatosensory receptive fields in the macaque putamen with corresponding visual receptive fields. *Exp. Brain Res.* 97, 96–109.
- Gross, D., Liu, Y. K., Flammia, S. T., Becker, S., and Eisert, J. (2010). Quantum state tomography via compressed sensing. *Phys. Rev. Lett.* 105, 150401. doi: 10.1103/PhysRevLett.105.150401
- Haider, B., Duque, A., Hasenstaub, A. R., and McCormick, D. A. (2006). Neocortical network activity *in vivo* is generated through a dynamic balance of excitation and inhibition. *J. Neurosci.* 26, 4535–4545. doi: 10.1523/JNEUROSCI.5297-05.2006
- Happe, F., Ronald, A., and Plomin, R. (2006). Time to give up on a single explanation for autism. *Nat. Neurosci.* 9, 1218–1220. doi: 10.1038/nn1770
- Haynes, J. D., Deichmann, R., and Rees, G. (2005). Eye-specific effects of binocular rivalry in the human lateral geniculate nucleus. *Nature* 438, 496–499. doi: 10.1038/nature04169
- He, Y., Chen, Z. J., and Evans, A. C. (2007). Small-world anatomical networks in the human brain revealed by cortical thickness from MRI. *Cereb. Cortex* 17, 2407–2419. doi: 10.1093/cercor/bhl149
- Heil, C., and Walnut, D. (1989). Continuous and discrete wavelet transforms. *SIAM Rev.* 31, 628–666.
- Herman, M. A., and Strohmer, T. (2009). High-resolution radar via compressed sensing. *Trans. Signal. Proc.* 57, 2275–2284. doi: 10.1109/TSP.2009.2014277
- Hodgkin, A. L., and Huxley, A. F. (1952). A quantitative description of membrane current and its application to conduction and excitation in nerve. *J. Physiol.* 117, 500–544.
- Holcombe, A. O., and Seizova-Cajic, T. (2008). Illusory motion reversals from unambiguous motion with visual, proprioceptive, and tactile stimuli. *Vision Res.* 48, 1743–1757. doi: 10.1016/j.visres.2008.05.019
- Hubel, D. (1995). *Eye, Brain, and Vision*. New York, NY: Henry Holt and Company.
- Hubel, D., and Wiesel, T. (1962). Receptive fields, binocular interaction and functional architecture of the cat's visual cortex. *J. Physiol.* 160, 106–154.
- Hubel, D. H., and Wiesel, T. N. (1960). Receptive fields of optic nerve fibres in the spider monkey. *J. Physiol.* 154, 572–580.
- Hupe, J. M., Joffo, L. M., and Pressnitzer, D. (2008). Bistability for audiovisual stimuli: perceptual decision is modality specific. *J. Vis.* 8, 1–15. doi: 10.1167/8.7.1
- Ingrasso, A., and Abbott, L. F. (2019). Training dynamically balanced excitatory-inhibitory networks. *PLoS ONE* 14, e0220547. doi: 10.1371/journal.pone.0220547
- Issa, N. P., Trachtenberg, J. T., Chapman, B., Zahs, K. R., and Stryker, M. P. (1999). The critical period for ocular dominance plasticity in the Ferret's visual cortex. *J. Neurosci.* 19, 6965–6978. doi: 10.1523/JNEUROSCI.19-16-06965.1999
- Jamain, S., Radyushkin, K., Hammerschmidt, K., Granon, S., Boretius, S., Varoquaux, F., et al. (2008). Reduced social interaction and ultrasonic communication in a mouse model of monogenic heritable autism. *Proc. Natl. Acad. Sci. U.S.A.* 105, 1710–1715. doi: 10.1073/pnas.0711555105
- Kilpatrick, Z. P., and Bressloff, P. C. (2010). Binocular rivalry in a competitive neural network with synaptic depression. *SIAM J. Appl. Dyn. Syst.* 9, 1303–1347. doi: 10.1137/100788872
- Klink, P. C., Brascamp, J. W., Blake, R., and van Wezel, R. J. (2010). Experience-driven plasticity in binocular vision. *Curr. Biol.* 20, 1464–1469. doi: 10.1016/j.cub.2010.06.057
- Knudsen, E. I., and Konishi, M. (1978). Center-surround organization of auditory receptive fields in the owl. *Science* 202, 778–780.
- Koch, C., Massimini, M., Boly, M., and Tononi, G. (2016). Neural correlates of consciousness: progress and problems. *Nat. Rev. Neurosci.* 17, 307–321. doi: 10.1038/nrn.2016.22
- Kormmeier, J., and Bach, M. (2012). Ambiguous figures—what happens in the brain when perception changes but not the stimulus. *Front. Hum. Neurosci.* 6, 51. doi: 10.3389/fnhum.2012.00051
- Kovacs, I., Papatomas, T. V., Yang, M., and Feher, A. (1996). When the brain changes its mind: interocular grouping during binocular rivalry. *Proc. Natl. Acad. Sci. U.S.A.* 93, 15508–15511.
- Krizhevsky, A., Sutskever, I., and Hinton, G. (2017). ImageNet classification with deep convolutional neural networks. *Commun. ACM* 60, 84–90. doi: 10.1145/3065386
- La Camera, G., Rauch, A., Thurbon, D., Luscher, H. R., Senn, W., and Fusi, S. (2006). Multiple time scales of temporal response in pyramidal and fast spiking cortical neurons. *J. Neurophysiol.* 96, 3448–3464. doi: 10.1152/jn.00453.2006
- Laing, C. R., and Chow, C. C. (2002). A spiking neuron model for binocular rivalry. *J. Comput. Neurosci.* 12, 39–53. doi: 10.1023/A:1014942129705
- Lansing, R. W. (1964). Electroencephalographic correlates of binocular rivalry in man. *Science* 146, 1325–1327.
- Lee, S. H., and Blake, R. (1999). Rival ideas about binocular rivalry. *Vision Res.* 39, 1447–1454.
- Lee, S. H., Blake, R., and Heeger, D. J. (2007). Hierarchy of cortical responses underlying binocular rivalry. *Nat. Neurosci.* 10, 1048–1054. doi: 10.1038/nn1939
- Leopold, D. A., and Logothetis, N. K. (1996). Activity changes in early visual cortex reflect monkeys' percepts during binocular rivalry. *Nature* 379, 549–553.
- Levelt, W. J. (1965). *On binocular rivalry* (Ph.D. thesis). Van Gorcum Assen.
- Li, H. H., Rankin, J., Rinzel, J., Carrasco, M., and Heeger, D. J. (2017). Attention model of binocular rivalry. *Proc. Natl. Acad. Sci. U.S.A.* 114, E6192–E6201. doi: 10.1073/pnas.1620475114
- Lim, S., and Goldman, M. S. (2014). Balanced cortical microcircuitry for spatial working memory based on corrective feedback control. *J. Neurosci.* 34, 6790–6806. doi: 10.1523/JNEUROSCI.4602-13.2014
- Liu, G. (2004). Local structural balance and functional interaction of excitatory and inhibitory synapses in hippocampal dendrites. *Nat. Neurosci.* 7, 373–379. doi: 10.1038/nn1206
- Logothetis, N. K., Leopold, D. A., and Sheinberg, D. L. (1996). What is rivaling during binocular rivalry? *Nature* 380, 621–624.
- London, M., Roth, A., Beeren, L., Hausser, M., and Latham, P. E. (2010). Sensitivity to perturbations *in vivo* implies high noise and suggests rate coding in cortex. *Nature* 466, 123–127. doi: 10.1038/nature09086
- Lustig, M., Donoho, D., and Pauly, J. M. (2007). Sparse MRI: the application of compressed sensing for rapid MR imaging. *Magn. Reson. Med.* 58, 1182–1195. doi: 10.1002/mrm.21391
- Markram, H., Lubke, J., Frotscher, M., Roth, A., and Sakmann, B. (1997). Physiology and anatomy of synaptic connections between thick tufted pyramidal neurones in the developing rat neocortex. *J. Physiol.* 500(Pt 2), 409–440.
- Mather, W., Bennett, M. R., Hasty, J., and Tsimring, L. S. (2009). Delay-induced degrade-and-fire oscillations in small genetic circuits. *Phys. Rev. Lett.* 102, 068105. doi: 10.1103/PhysRevLett.102.068105
- Matsuoka, K. (1984). The dynamic model of binocular rivalry. *Biol. Cybern.* 49, 201–208.
- McKinstry, J. L., Fleischer, J. G., Chen, Y., Gall, W. E., and Edelman, G. M. (2016). Imagery may arise from associations formed through sensory experience: a network of spiking neurons controlling a robot learns visual sequences in order to perform a mental rotation task. *PLoS ONE* 11, e0162155. doi: 10.1371/journal.pone.0162155
- McLaughlin, D., Shapley, R., Shelley, M., and Wielaard, J. (2000). A neuronal network model of macaque primary visual cortex (V1): Orientation selectivity and dynamics in the input layer 4C α . *Proc. Natl. Acad. Sci. U.S.A.* 97, 8087–8092. doi: 10.1073/pnas.110135097
- Mentch, J., Spiegel, A., Ricciardi, C., and Robertson, C. E. (2019). GABAergic inhibition gates perceptual awareness during binocular rivalry. *J. Neurosci.* 39, 8398–8407. doi: 10.1523/JNEUROSCI.0836-19.2019
- Miller, S. M. (2013). *The Constitution of Visual Consciousness: Lessons From Binocular Rivalry*, Vol. 90. Amsterdam: John Benjamins Publishing.
- Miller, S. M., Hansell, N. K., Ngo, T. T., Liu, G. B., Pettigrew, J. D., Martin, N. G., et al. (2010). Genetic contribution to individual variation in binocular rivalry rate. *Proc. Natl. Acad. Sci. U.S.A.* 107, 2664–2668. doi: 10.1073/pnas.0912149107
- Miura, K., Tsubo, Y., Okada, M., and Fukai, T. (2007). Balanced excitatory and inhibitory inputs to cortical neurons decouple firing irregularity from rate modulations. *J. Neurosci.* 27, 13802–13812. doi: 10.1523/JNEUROSCI.2452-07.2007
- Mori, K., Nagao, H., and Yoshihara, Y. (1999). The olfactory bulb: coding and processing of odor molecule information. *Science* 286, 711–715.
- Nawrot, M., and Blake, R. (1989). Neural integration of information specifying structure from stereopsis and motion. *Science* 244, 716–718.
- Nelson, J. I. (1975). Globality and stereoscopic fusion in binocular vision. *J. Theor. Biol.* 49, 1–88.
- Nelson, S. B., and Valakh, V. (2015). Excitatory/inhibitory balance and circuit homeostasis in autism spectrum disorders. *Neuron* 87, 684–698. doi: 10.1016/j.neuron.2015.07.033

- Nguyen, Q. A., Rinzel, J., and Curtu, R. (2020). Buildup and bistability in auditory streaming as an evidence accumulation process with saturation. *PLoS Comput. Biol.* 16, e1008152. doi: 10.1371/journal.pcbi.1008152
- Noor, A., Serpedin, E., Nounou, M., and Nounou, H. (2013). Reverse engineering sparse gene regulatory networks using cubature kalman filter and compressed sensing. *Adv. Bioinformatics* 2013, 205763. doi: 10.1155/2013/205763
- Pearson, J., and Brascamp, J. (2008). Sensory memory for ambiguous vision. *Trends Cogn. Sci.* 12, 334–341. doi: 10.1016/j.tics.2008.05.006
- Platonov, A., and Goossens, J. (2013). Influence of contrast and coherence on the temporal dynamics of binocular motion rivalry. *PLoS ONE* 8, e71931. doi: 10.1371/journal.pone.0071931
- Polonsky, A., Blake, R., Braun, J., and Heeger, D. J. (2000). Neuronal activity in human primary visual cortex correlates with perception during binocular rivalry. *Nat. Neurosci.* 3, 1153–1159. doi: 10.1038/80676
- Rangan, A. V., and Cai, D. (2006). Maximum-entropy closures for kinetic theories of neuronal network dynamics. *Phys. Rev. Lett.* 96, 178101. doi: 10.1103/PhysRevLett.96.178101
- Rauch, A., La Camera, G., Luscher, H.-R., Senn, W., and Fusi, S. (2003). Neocortical pyramidal cells respond as integrate-and-fire neurons to *in vivo*-like input currents. *J. Neurophysiol.* 90, 1598–1612. doi: 10.1152/jn.00293.2003
- Robertson, C. E., Kravitz, D. J., Freyberg, J., Baron-Cohen, S., and Baker, C. I. (2013). Slower rate of binocular rivalry in autism. *J. Neurosci.* 33, 16983–16991. doi: 10.1523/JNEUROSCI.0448-13.2013
- Robertson, C. E., Ratai, E. M., and Kanwisher, N. (2016). Reduced GABAergic action in the autistic brain. *Curr. Biol.* 26, 80–85. doi: 10.1016/j.cub.2015.11.019
- Rosenberg, A., Patterson, J. S., and Angelaki, D. E. (2015). A computational perspective on autism. *Proc. Natl. Acad. Sci. U.S.A.* 112, 9158–9165. doi: 10.1073/pnas.1510583112
- Rozell, C. J., Johnson, D. H., Baraniuk, R. G., and Olshausen, B. A. (2008). Sparse coding via thresholding and local competition in neural circuits. *Neural Comput.* 20, 2526–2563. doi: 10.1162/neco.2008.03-07-486
- Rubenstein, J. L., and Merzenich, M. M. (2003). Model of autism: increased ratio of excitation/inhibition in key neural systems. *Genes Brain Behav.* 2, 255–267. doi: 10.1034/j.1601-183X.2003.00037.x
- Rutishauser, U., Douglas, R. J., and Slotine, J. J. (2011). Collective stability of networks of winner-take-all circuits. *Neural Comput.* 23, 735–773. doi: 10.1162/NECO_a_00091
- Said, C. P., and Heeger, D. J. (2013). A model of binocular rivalry and cross-orientation suppression. *PLoS Comput. Biol.* 9, e1002991. doi: 10.1371/journal.pcbi.1002991
- Sceniak, M. P., Ringach, D. L., Hawken, M. J., and Shapley, R. (1999). Contrast's effect on spatial summation by macaque V1 neurons. *Nat. Neurosci.* 2, 733–739.
- Seely, J., and Chow, C. C. (2011). Role of mutual inhibition in binocular rivalry. *J. Neurophysiol.* 106, 2136–2150. doi: 10.1152/jn.00228.2011
- Sengpiel, F., Blakemore, C., Kind, P. C., and Harrad, R. (1994). Interocular suppression in the visual cortex of strabismic cats. *J. Neurosci.* 14(11 Pt 2), 6855–6871.
- Shannon, C. E. (1949). Communication in the presence of noise. *Proc. IRE* 37, 10–21.
- Shannon, R. W., Patrick, C. J., Jiang, Y., Bernat, E., and He, S. (2011). Genes contribute to the switching dynamics of bistable perception. *J. Vis.* 11, 17. doi: 10.1167/11.3.8
- Shapiro, A., Moreno-Bote, R., Rubin, N., and Rinzel, J. (2009). Balance between noise and adaptation in competition models of perceptual bistability. *J. Comput. Neurosci.* 27, 37–54. doi: 10.1007/s10827-008-0125-3
- Spear, P. D., Kim, C. B., Ahmad, A., and Tom, B. W. (1996). Relationship between numbers of retinal ganglion cells and lateral geniculate neurons in the rhesus monkey. *Vis. Neurosci.* 13, 199–203.
- Spiegel, A., Mentch, J., Haskins, A. J., and Robertson, C. E. (2019). Slower binocular rivalry in the autistic brain. *Curr. Biol.* 29, 2948–2953. doi: 10.1016/j.cub.2019.07.026
- Sterzer, P., Kleinschmidt, A., and Rees, G. (2009). The neural bases of multistable perception. *Trends Cogn. Sci.* 13, 310–318. doi: 10.1016/j.tics.2009.04.006
- Stettler, D. D., Das, A., Bennett, J., and Gilbert, C. D. (2002). Lateral connectivity and contextual interactions in macaque primary visual cortex. *Neuron* 36, 739–750. doi: 10.1016/S0896-6273(02)01029-2
- Suzuki, S., and Grabowecy, M. (2002). Evidence for perceptual “trapping” and adaptation in multistable binocular rivalry. *Neuron* 36, 143–157. doi: 10.1016/S0896-6273(02)00934-0
- Tamamaki, N., and Tomioka, R. (2010). Long-range GABAergic connections distributed throughout the neocortex and their possible function. *Front. Neurosci.* 4, 202. doi: 10.3389/fnins.2010.00202
- Tao, L., Cai, D., McLaughlin, D., Shelley, M., and Shapley, R. (2006). Orientation selectivity in visual cortex by fluctuation-controlled criticality. *Proc. Natl. Acad. Sci. U.S.A.* 103, 12911–12916. doi: 10.1073/pnas.0605415103
- Tong, F. (2001). Competing theories of binocular rivalry: a possible resolution. *Brain Mind* 2, 55–83. doi: 10.1023/A:1017942718744
- Tong, F., and Engel, S. A. (2001). Interocular rivalry revealed in the human cortical blind-spot representation. *Nature* 411, 195–199. doi: 10.1038/35075583
- Tong, F., Meng, M., and Blake, R. (2006). Neural bases of binocular rivalry. *Trends Cogn. Sci.* 10, 502–511. doi: 10.1016/j.tics.2006.09.003
- Tropp, J. A., and Gilbert, A. C. (2007). Signal recovery from random measurements via orthogonal matching pursuit. *IEEE Trans. Inform. Theory* 53, 4655–4666. doi: 10.1109/TIT.2007.909108
- Troyer, T. W., and Miller, K. D. (1997). Physiological gain leads to high ISI variability in a simple model of a cortical regular spiking cell. *Neural Comput.* 9, 971–983.
- van Ee, R. (2009). Stochastic variations in sensory awareness are driven by noisy neuronal adaptation: evidence from serial correlations in perceptual bistability. *J. Opt. Soc. Am. A Opt. Image Sci. Vis.* 26, 2612–2622. doi: 10.1364/JOSAA.26.002612
- van Vreeswijk, C., and Sompolinsky, H. (1996). Chaos in neuronal networks with balanced excitatory and inhibitory activity. *Science* 274, 1724–1726.
- van Vreeswijk, C., and Sompolinsky, H. (1998). Chaotic balanced state in a model of cortical circuits. *Neural Comput.* 15, 1321–1371.
- Vattikuti, S., and Chow, C. C. (2010). A computational model for cerebral cortical dysfunction in autism spectrum disorders. *Biol. Psychiatry* 67, 672–678. doi: 10.1016/j.biopsych.2009.09.008
- Vattikuti, S., Thangaraj, P., Xie, H. W., Gotts, S. J., Martin, A., and Chow, C. C. (2016). Canonical cortical circuit model explains rivalry, intermittent rivalry, and rivalry memory. *PLoS Comput. Biol.* 12, e1004903. doi: 10.1371/journal.pcbi.1004903
- Wang, W.-X., Lai, Y.-C., Grebogi, C., and Ye, J. (2011). Network reconstruction based on evolutionary-game data via compressive sensing. *Phys. Rev. X* 1, 021021. doi: 10.1103/PhysRevX.1.021021
- Wang, Z., Dai, W., and McLaughlin, D. W. (2020). Ring models of binocular rivalry and fusion. *J. Comput. Neurosci.* 48, 193–211. doi: 10.1007/s10827-020-00744-7
- Welker, C. (1976). Receptive fields of barrels in the somatosensory neocortex of the rat. *J. Comp. Neurol.* 166, 173–189.
- Wheatstone, C. (1838). XVIII. contributions to the physiology of vision.—Part the first. on some remarkable, and hitherto unobserved, phenomena of binocular vision. *Philos. Trans. R. Soc. Lond.* 128, 371–394.
- Wiesel, T. N. (1960). Receptive fields of ganglion cells in the cat's retina. *J. Physiol.* 153, 583–594.
- Wilke, M., Logothetis, N. K., and Leopold, D. A. (2003). Generalized flash suppression of salient visual targets. *Neuron* 39, 1043–1052. doi: 10.1016/j.neuron.2003.08.003
- Wilson, D. A. (2001). Receptive fields in the rat piriform cortex. *Chem. Senses* 26, 577–584. doi: 10.1093/chemse/26.5.577
- Wilson, H. R. (2003). Computational evidence for a rivalry hierarchy in vision. *Proc. Natl. Acad. Sci. U.S.A.* 100, 14499–14503. doi: 10.1073/pnas.2333622100
- Wilson, H. R. (2013). “Binocular rivalry: cooperation, competition, and decisions,” in *The Constitution of Visual Consciousness*, ed S. M. Miller (Amsterdam: John Benjamins), 281–304.
- Wilson, H. R. (2017). Binocular contrast, stereopsis, and rivalry: toward a dynamical synthesis. *Vision Res.* 140, 89–95. doi: 10.1016/j.visres.2017.07.016
- Wilson, H. R., Blake, R., and Lee, S. H. (2001). Dynamics of travelling waves in visual perception. *Nature* 412, 907–910. doi: 10.1038/35091066
- Xu, H., Han, C., Chen, M., Li, P., Zhu, S., Fang, Y., et al. (2016). Rivalry-like neural activity in primary visual cortex in anesthetized monkeys. *J. Neurosci.* 36, 3231–3242. doi: 10.1523/JNEUROSCI.3660-15.2016
- Xue, M., Atallah, B. V., and Scanziani, M. (2014). Equalizing excitation-inhibition ratios across visual cortical neurons. *Nature* 511, 596–600. doi: 10.1038/nature13321
- Yamada, W., Koch, C., and Adams, P. (1989). “Multiple channels and calcium dynamics,” in *Methods in Neuronal Modeling: From Synapses to Networks* (MIT Press), 97–133.
- Zhang, P., Jamison, K., Engel, S., He, B., and He, S. (2011). Binocular rivalry requires visual attention. *Neuron* 71, 362–369. doi: 10.1016/j.neuron.2011.05.035
- Zhou, W., and Chen, D. (2009). Binocular rivalry between the nostrils and in the cortex. *Curr. Biol.* 19, 1561–1565. doi: 10.1016/j.cub.2009.07.052