Reviewed Preprint
v1 • August 11, 2026
Not revised

Reviewed Preprint
v2 • October 26, 2026
Revised by authors

Hierarchical Neural Circuit Theory of Normalization and Inter-areal Communication

Asit Pal[1,2], **Shivang Rawat**[2,3], **David J Heeger**[4,5], **Stefano Martiniani**[1,2,3,5]

[1]Simons Center for Computational Physical Chemistry, Dept. of Chemistry, New York University, New York, United States • [2]Center for Soft Matter Research, Dept. of Physics, New York University, New York, United States • [3]Courant Institute of Mathematical Sciences, New York University, New York, United States • [4]Dept. of Psychology, New York University, New York, United States • [5]Center for Neural Science, New York University, New York, United States

For correspondence:
asit.pal@nyu.edu
david.heeger@nyu.edu
stefano.martiniani@nyu.edu

Competing interests: David Heeger is co-founder and chief scientific officer of The Sequel Institute, which has licensed IP from NYU related to ORGaNICs. Shivang Rawat is currently a full-time employee at The Sequel Institute. Stefano Martiniani is a consultant for The Sequel Institute.

Funding: See page 29

Reviewing editor: Audrey Sederberg, Georgia Institute of Technology, United States

© 2026, Pal et al. This article is distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use and redistribution provided that the original author and source are credited.

eLife Assessment

This **important** study extends a model of cortical normalization (ORGaNICs) to interacting cortical areas and shows that communication through coherence and communication subspaces can arise from a single set of dynamics. The evidence is **convincing**, showing analytically that contrast-dependent gamma dynamics and a low-dimensional inter-areal communication subspace arise from one parameter set. The work will interest neuroscientists and theorists concerned with inter-areal communication, cortical oscillations, and divisive normalization.

https://doi.org/10.7554/eLife.111297.2.sa4

Abstract

The primate brain exhibits a hierarchical, modular architecture with conserved microcircuits executing canonical computations across reciprocally connected cortical areas. We present a hierarchical neural circuit theory with feedback connections that dynamically implements divisive normalization across its hierarchy. In a two-stage instantiation (V1 ↔ V2), increasing feedback from V2 to V1 amplifies responses in both areas. We analytically derive power spectra (V1) and coherence spectra (V1-V2) as functions of frequency (f), and compare them with experimental observations: peaks in both spectra shift to higher frequencies with increased stimulus contrast, and power decays as $1/f^4$ at high frequencies. The closed-form spectra are validated against direct stochastic simulation of the full nonlinear circuit across contrasts. The theory further predicts distinctive spectral signatures of feedback and input gain modulation. Crucially, the theory offers a unified view of inter-areal communication, defined as the ability to linearly predict neural activity in one brain area from neural activity in another, with emergent features consistent with empirical observations of both communication subspaces and inter-areal coherence. It admits a low-dimensional communication subspace, where inter-areal communication is lower-dimensional than within-area communication. It further predicts that: i) increasing feedback strength enhances inter-areal communication and diminishes within-area communication, without altering the subspace dimensionality; ii) communication is strongest, and the subspace dimensionality lowest, at the frequencies at which V1-V2 coherence peaks; iii) normalization reduces the subspace dimensionality. Finally, a three-area (V1 ↔ V4 and V1 ↔ V5) instantiation of the theory demonstrates that differential feedback from higher to lower cortical areas dictates their dynamic functional connectivity. Altogether, our theory provides a robust and analytically tractable framework for generating experimentally-testable predictions about normalization, inter-areal communication, and functional connectivity.

Introduction

The primate brain is a complex hierarchical network of interconnected cortical areas that perform computations essential for sensorimotor processing and cognitive functions. Among these, certain computations are repeated across cortical areas and are thus said to be "canonical computations". Divisive normalization [1] is one such fundamental canonical principle, playing a central role in regulating neural responses across sensory and higher cortical areas.

Feedback connections are ubiquitous throughout the primate brain, yet their precise computational role remains poorly understood. A wide range of functions has been attributed to feedback, including attentional modulation [2], inter-areal communication [3], learning [4], and generative modeling [5]. While divisive normalization has been extensively studied in feedforward recurrent networks [6–9], its implementation and effects in hierarchical networks with feedback remain unexplored.

Effective communication between brain areas is essential for complex cognitive tasks, yet the precise mechanisms of communication remain an active area of investigation. One prominent hypothesis, *communication through coherence* (CTC) [10], proposes that synchronized oscillations (coherence) are crucial for effective inter-areal communication. Indeed, coherence between brain regions has been widely observed across cognitive, perceptual, and behavioral tasks [11, 12]. However, the causal nature of the relationship remains a key unresolved question. It is currently debated whether coherence is a fundamental mechanism enabling inter-areal communication or rather an epiphenomenon emerging from underlying interactions [13, 14]. An alternative hypothesis suggests that brain areas communicate through a low-dimensional *communication subspace* (CS) embedded within the high-dimensional neural activity space [15]. This concept, also identified in artificial neural networks [16], posits that only activity within these shared subspaces between the source and target areas facilitates communication. Recent modeling work has begun to examine how communication subspaces and the directionality of inter-areal interactions arise from circuit dynamics and connectivity [17–20]. Despite strong empirical evidence for both coherence-based and subspace-based communication, a unified, analytically tractable framework that jointly accounts for coherence and communication subspaces and explaining how they are dynamically modulated by factors such as feedback and stimulus contrast is currently lacking.

To address these gaps, we introduce a hierarchical recurrent neural circuit theory (Fig. 1) built upon two *a priori* assumptions: i) a hierarchy of brain areas with feedforward, recurrent, and feedback connections, and ii) recurrently implemented normalization within each brain area. A concrete model also requires choices about the connectivity, the normalization pool, and the noise (Methods); their influence on the results is examined in Suppl. Fig. S4 and in the *SI Appendix* ("Robustness with respect to model parameters"). Within this framework, we derive analytical expressions for inter-areal coherence and communication subspaces, offering experimentally testable predictions (see Discussion, "Experimental predictions of the model"). Given these choices, and a single set of parameters, the theory gives rise to a rich set of emergent properties that account for a wide spectrum of neurophysiological phenomena, including modulation of contrast response functions, oscillatory dynamics evident in power and coherence spectra, the emergence of communication subspaces, and the dynamic control of functional connectivity. A central contribution of this work is a unified view of interareal communication. The theory simultaneously predicts empirical observations of both communication subspaces and inter-areal coherence, and further reveals that coherence actively shapes the communication subspace. In doing so, our framework reconciles two hypotheses that have hitherto been treated as alternatives, and shows that coherence and subspace structure are two emergent properties of the same underlying circuit.

Finally, we emphasize that throughout our analysis, we neither adjust model parameters (such as time constants, weight matrices, and noise amplitudes) to reproduce specific experimental trends nor fit data when making direct comparisons. All results presented were obtained using a *single*

set of model parameters. The only exception is Suppl. Fig. S7, in which the parameters are fitted to measurements of LFP power spectra. Additionally, we have verified that the key results hold across a wide range of parameters; for further details, see the *SI Appendix*.

Results

Theory

Initially conceived to explain the responses in the primary visual cortex (V1) [1], the normalization model has emerged as a canonical computation across sensory and cognitive domains [21], including higher level vision, olfaction, audition, somatosensation, attention, multi-sensory integration, working memory, and decision-making. For citations see [6]. At its core, the normalization model proposes that the response of an individual neuron is scaled by the collective activity of its neighboring neurons (viz., the normalization pool), analogous to normalizing the length of a vector.

In addition, normalization decorrelates population responses to reduce redundancy [22, 23],maintains network stability in spite of recurrent amplification [8, 9], and generates long timescales without fine-tuning [24]. It is a computational-level description of neural responses that can be implemented using various mechanisms. Experimental evidence suggests that normalization operates via recurrent amplification [25, 26], amplifying weak inputs more than strong inputs, consonant with evidence that cortical circuits are dominated by recurrent excitation [27, 28].

Oscillatory **R**ecurrent **G**ated **N**eural **I**ntegrator **C**ircuits (ORGaNICs) is a theoretical framework that dynamically implements divisive normalization by modulating recurrent amplification [6, 7]. Previous work on normalization models (including previous work on ORGaNICs) has focused on a single neural circuit within a brain region, such as V1, ignoring the abundant top-down feedback connections. Here, we develop a new hierarchical ORGaNICs theory with feedback connections in addition to feedforward and recurrent connections.

The hierarchical ORGaNICs theory describes neuronal population responses as dynamic processes evolving in a recurrent circuit, modeled by nonlinear differential equations (see Fig. 1 for illustration and Methods for equations). The circuit comprises distinct cell types: principal cells (**y**), excitatory modulator cells (**u**), and inhibitory modulator cells (**a, q**) (n.b., “modulator” refers to a multiplicative computation, not necessarily implemented with neuro-modulators). Output firing rates of principal cells depend on three terms: 1) feedforward input drive modulated (viz., scaled) by input gain; 2) feedback drive modulated by feedback gain; 3) the sum of recurrent drive and scaled feedback drive, collectively modulated by recurrent gain (see equation at the top of Fig. 1). Whereas in Wilson–Cowan-type excitatory–inhibitory models apply the same form of rate equation to each subpopulation of neurons, with different signs and weights, the four populations here obey structurally different equations, each implementing a distinct operation (Methods, Eq. 1). Specifically, principal cells **y** sum the gain-modulated input, recurrent, and feedback drives (with a squaring output nonlinearity); excitatory modulator cells **u** accumulate the normalization signal, a weighted sum of principal-cell responses (square-root output nonlinearity); inhibitory modulator cells **a** implement the division, i.e., the multiplicative gain control of the recurrent and feedback drives; and inhibitory cells **q** track principal-cell activity and provide the sign inversion required by the feedback term in the dynamics of **a** (*SI Appendix*, Eqs. S21 and S22). This theory is particularly well-suited to analytical treatment, e.g., to derive closed-form solutions for the power spectral density, coherence, and communication subspaces (see Methods and Appendix).

Each neuron receives *input drive* defined as a weighted sum of responses from neurons at lower stages of the hierarchy. For example, LGN responses provide input drive to V1 ($\mathbf{z_1}$ in Fig.1), and V1 responses similarly drive V2 (solid arrow from $\mathbf{y_1}$ to $\mathbf{y_2}$ in Fig.1, representing principal neuron responses in each area). This weighting is specified by an embedding matrix of effective weights, which determine each neuron’s stimulus preferences, such as orientation selectivity (Fig.1, V1 tuning curves), analogous to a standard receptive field model (see Methods). Throughout, the weight matrices describe effective connectivity rather than anatomical synapses: negative weights

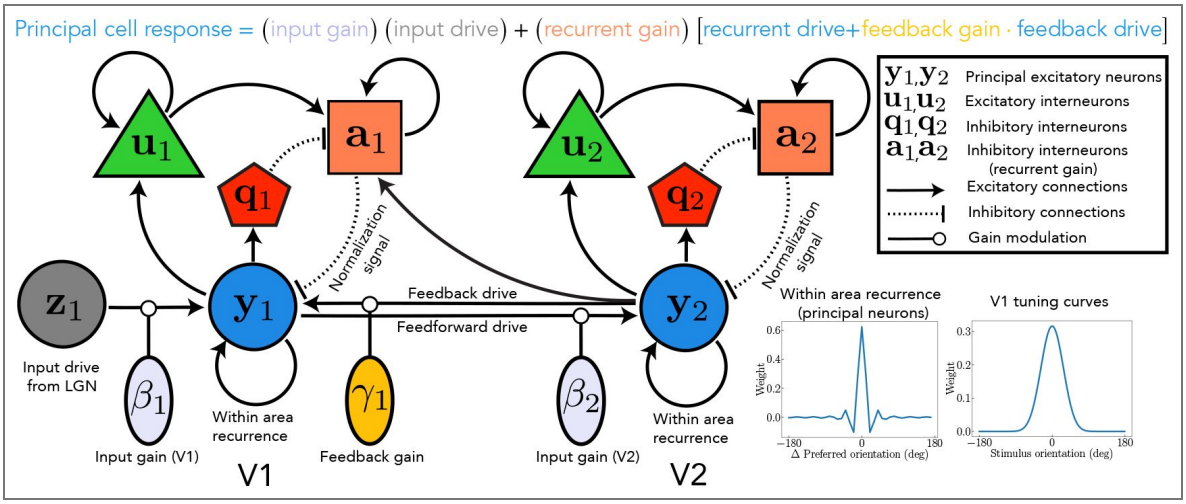


Figure 1. Schematic of the two-stage recurrent circuit model.

Solid and dashed lines represent net excitatory and net inhibitory (effective) connections, respectively. Principal neurons are denoted by **y**, modulatory excitatory neurons by **u**, and modulatory inhibitory neurons by **a** and **q**. Subscripts 1 and 2 designate neurons in visual areas V1 and V2, respectively. $\mathbf{z}_1$ denotes the input drive from the LGN to V1, a weighted sum of the responses of LGN neurons. Parameters β_1 and γ_1 modulate the input and feedback gain to V1, respectively.

may, for example, be implemented with disynaptic inhibition through interneurons that are not modeled explicitly (see Discussion). Input gain is determined independently by input modulator cells (β_1 and β_2 in Fig.1), whose gain may, in principle, depend on both afferent input and local population activity (e.g., responses of both LGN cells and V1 cells may contribute to the input gain in V1).

Each neuron also receives *recurrent drive* arising from weighted interactions among principal cells in the same brain area (Fig. 1, curved arrows). Recurrent connectivity follows a center-surround architecture, with local excitatory and distal inhibitory connections (Fig. 1, recurrent synaptic weights). The gain of this recurrent drive is dynamically controlled by recurrent modulator cells ($\mathbf{a}_1$ and $\mathbf{a}_2$), which combine excitatory ($\mathbf{u}_1$, $\mathbf{u}_2$) and inhibitory signals ($\mathbf{q}_1$, $\mathbf{q}_2$) from local interneuron populations as well as feedback from higher cortical areas (solid arrow from $\mathbf{y}_2$ to $\mathbf{a}_1$ in Fig. 1). Divisive normalization is achieved dynamically through the activity of these modulator cells, which carry the "normalization signal".

Finally, each area receives *feedback drive*, a weighted sum of responses from neurons at higher stages of the hierarchy (e.g., V2 → V1; Fig. 1). The feedback drive is specified by a feedback weight matrix, which, for simplicity, we take to be the transpose of the corresponding feedforward matrix from V1 to V2 [5]. The strength of this feedback is controlled by a feedback gain parameter (γ_1 in Fig. 1). More generally, feedback strength may be modulated either by changing the feedback gain or by changing the gain of the neural responses in the higher cortical areas providing feedback (see Discussion). Thus, the feedback gain may be a fixed property of the circuit or a dynamically regulated quantity that depends on both the inputs and outputs of the corresponding area. Importantly, even if the feedback gain is fixed, it may still be possible to manipulate the feedback strength experimentally to test predictions of the theory (elaborated in what follows).

The two-stage recurrent model with feedback implements divisive normalization simultaneously across multiple cortical areas. The implementation is exact in a special case: when the gains are set to $\gamma = \beta = 1$, and the recurrent weight matrix is the identity matrix (i.e., each neuron recurrently amplifies only itself), the steady-state responses of the model coincide precisely with the analytical steadystate solutions of the normalization equation (Methods, Eq. 2). For the center-surround recurrent weight matrix used in all results reported here, the circuit very closely approximates normalization (see Methods and Ref. [9]). The exact fixed point is obtained numerically by integrating the deterministic dynamics to steady state, and all spectra and subspaces are then computed in closed form by linearizing about this fixed point (see Methods).

The core mechanism in brief

Noise-driven fluctuations around the normalized fixed point are characterized by the cross-spectral density matrix $\boldsymbol{S}(\omega)$ (Methods, Eq. 6). Its diagonal entries determine the power spectra of individual neurons, while its normalized off-diagonal entries determine inter-areal coherence (Eq. 7). Integrating $\boldsymbol{S}(\omega)$ over frequency yields the covariance matrix, and its frequency-resolved counterpart provides the covariance structure from which the communication subspace is obtained by reduced-rank regression (Methods). Thus, power, coherence, and communication-subspace structure are complementary readouts of the same underlying fluctuation dynamics.

Recurrent normalization shapes these dynamics in two related ways. First, recurrent gain control modifies the circuit dynamics around the fixed point, giving rise to contrast-dependent resonances that appear as peaks in both power and coherence. Second, competition within the normalization pool preferentially suppresses weakly driven neurons, producing a soft winner-take-all effect that reduces their contribution to the cross-covariance and thereby lowers the effective dimensionality of the communication subspace, particularly near the resonant frequencies. Removing normalization abolishes the resonant peaks in power and coherence and produces a higher-dimensional communication subspace (Suppl. Figs. S4e–h and S5).

Contrast response functions

Both V1 (dashed) and V2 (solid) exhibit the characteristic saturating, sigmoidal dependence on contrast (As shown in Fig. 2a, b). The model predicts a lower semisaturation constant (the contrast at which the firing rate reaches half-saturation) and a steeper slope for cortical area V2 compared to V1. This prediction is commensurate with experimental observations in the literature that used optimal stimulus sizes (Fig. 2b) [29, 30]. Similar hierarchical trends, specifically an increase in slope and a decrease in the semisaturation constant, have also been observed across other brain areas like the LGN, V1, and MT with progression up the visual hierarchy [32].

Increasing either the feedback gain, γ_1, or the V1 input gain, β_1 (with the V2 input gain β_2 held at its baseline value), enhances neural responses across cortical areas (Figs. 3a,b). Importantly, this response enhancement is evident primarily as a change in contrast gain, i.e., a shift of the contrast-response functions on the log contrast axis rather than a change in response amplitude. Increasing the V2 input gain, β_2, leaves V1 responses unchanged and enhances only V2 responses (Suppl. Fig. S8c).

V1 power spectra

Power spectrum analysis of electroencephalographic (EEG) and local field potential (LFP) signals has been used extensively to identify and quantify oscillatory activity in various frequency bands (e.g., alpha, beta, gamma, theta), hypothesized to be linked to various cognitive functions, behavioral states, and neurological conditions [33]. Since our model has a known fixed point solution, we calculate the power and coherence spectra analytically, given a model of the noise in the circuit (see Methods). The analytical spectra rest on a linearization of the dynamics about the fixed point; we verified them against direct numerical integration of the full nonlinear stochastic circuit at every contrast (Suppl. Fig. S6).

We adopted an analytically tractable noise model that adds low-pass filtered synaptic noise to the membrane potential and multiplicative Gaussian white noise to the firing rate dynamics. Furthermore, we explicitly accounted for long-range correlated noise in LFP measurements. This type of noise, propagating through brain tissue, affects LFPs with negligible impact on spiking activity and firing rates [34]. Further details are provided in the *SI Appendix* ("Intrinsic neural circuit noise" and "Extrinsic noise in the local field potential").

Our theory reproduces the main qualitative features of the V1 power spectra across varying contrast levels. These theoretical V1 predictions (Fig. 2c) are compared with experimental spectra from macaque V1 (Fig. 2d) under identical contrast conditions [31]. With the unfitted baseline parameters, the model captures three qualitative trends: 1) the power spectrum of the noise is shaped by recurrent normalization to exhibit a resonant peak in the gamma-frequency range (≈ 40-60 Hz); 2) the peak frequency shifts right-ward with increasing contrast; and 3) the peak power decreases after reaching a maximum around 40% contrast. A quantitative comparison, in which the model parameters are fitted to the measured spectra, is provided in Suppl. Fig. S7a (*SI Appendix*, "Quantitative fit to experimental power and coherence spectra"). The oscillatory activity in the gamma-frequency range depends on normalization; removing normalization eliminates the peaks (see Suppl. Fig. S4e), commensurate with experimental observations that gamma oscillations are linked to normalization [35, 36]. The peak frequencies depend on the values of the time constants; larger time constants would shift the peaks to lower frequencies (see Discussion: Limitations).

The theory also reproduces the steep fall-off in power reported at higher frequencies [37], where power approximately scales as $1/f^4$ with increasing frequency (f) (Fig. 2c, inset). This occurs because the synaptic noise added to the membrane potentials is first low-pass filtered by the synapse's time constant, and then by recurrent processing in the circuit, which also acts as a low-pass filter (each contributing a factor of $1/f^2$ to the power spectrum).

The theory offers a novel, experimentally testable prediction that modulating feedback, feedforward, and input gains elicits distinct changes in the power spectra. Figures 3c-f illustrate these distinct effects on the normalized power spectra when feedback gain (left panels) or input

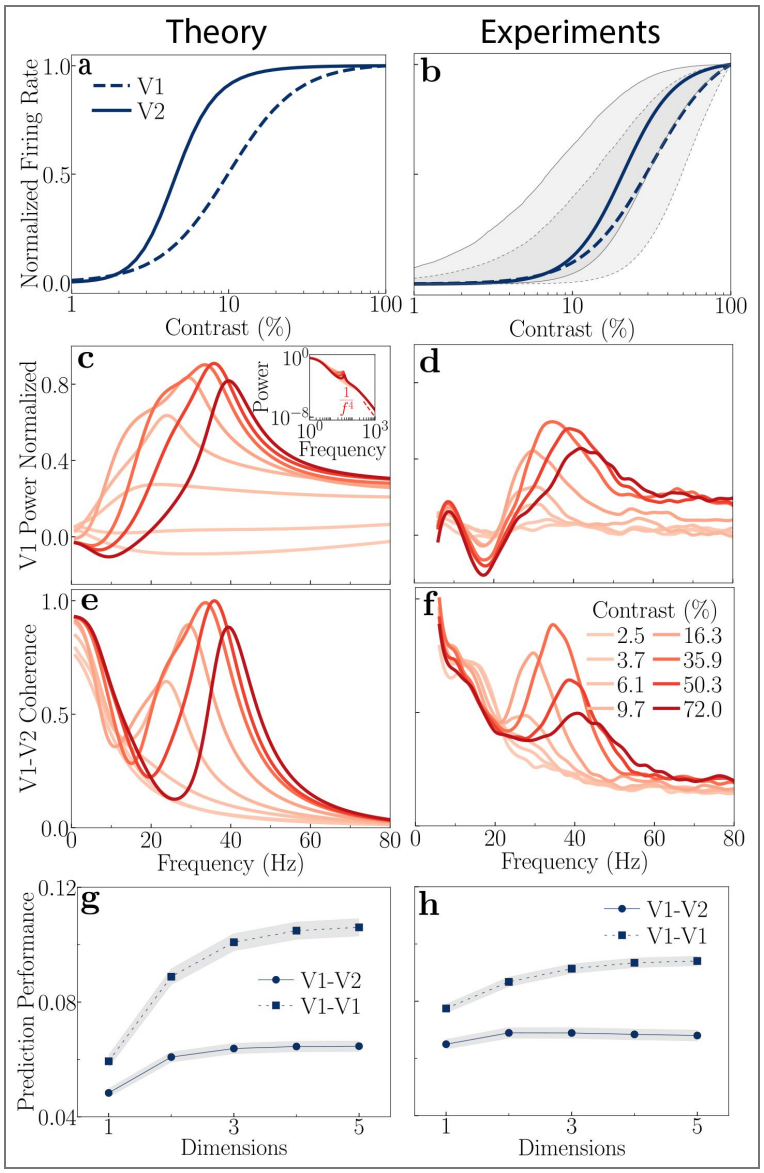


Figure 2. Comparison of theoretical predictions (left column) and experimental observations (right column) in V1 and V2.

All theoretical predictions are generated using the same baseline parameters (Table S1). The analytical spectra in **c** and **e** are validated against direct stochastic simulation of the full nonlinear circuit in Suppl. Fig. S6, and a quantitative comparison in which the model parameters are fitted to experimental measurements of the power and coherence spectra is given in Suppl. Fig. S7. **a**, Theoretical mean firing rates as a function of stimulus contrast (V1: dashed line, V2: solid line). **b**, Experimental mean firing rates of V1 and V2 (data from [29] and [30], replotted for comparison). The shaded area with a solid border indicates the 25th to 75th percentile range for V1, and the one with the dashed border indicates the same for V2. **c**, Theoretical V1 power spectra at various stimulus contrast levels. Power spectra were normalized using the equation $\frac{(\text{power}-\text{baseline})}{(\text{power}+\text{baseline})}$, where baseline power is taken to be spontaneous activity at 0% contrast. The peak power shifts towards higher frequency with increasing contrast. The inset shows the $\frac{1}{f^4}$ power law decay at high frequency. **d**, Experimental power spectra from macaque V1 (data from [31], replotted for comparison). **e**, Theoretical coherence spectra (normalized) between maximally firing neurons in V1 and V2. The peak coherence shifts towards higher frequency with increasing contrast. **f**, Experimental coherence spectra from macaque V1-V2 (data from [31], replotted for comparison). **g**, Theoretical communication subspaces, prediction performance as a function of dimensionality for inter-area (circles) and within-area (squares) communication. Simulation results indicate that the inter-area communication subspace is lower dimensional than the within-area communication. **h**, Experimental communication subspaces from macaque V1-V2 (data from [15], replotted for comparison). Th**5**e plotted prediction performance in both panels **g** and **h** is an average across different subsets of the source and target neural populations, and the shaded areas represent the standard error of the mean (SEM).

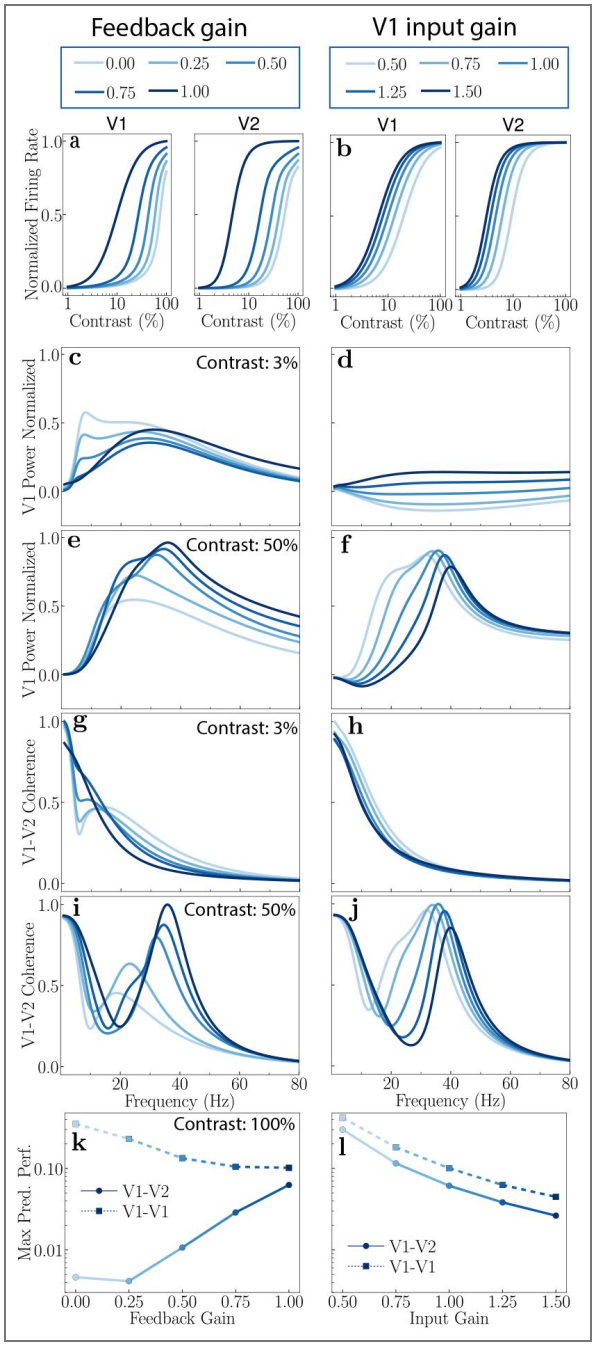


Figure 3. Theoretical predictions: Modulating feedback (left) and V1 input gain (right).

a,**b**, Firing rates as a function of contrast. Increasing either feedback gain or V1 input gain enhances neural responses, with feedback gain showing a greater impact in the higher cortical area V2. **c**,**d**, V1 power spectra: 3% contrast. An alpha peak is observed for low feedback gain, but the peak diminishes with increasing feedback gain. The alpha peak is absent with V1 input gain changes. **e**,**f**, V1 power spectra: 50% contrast. A consistent gamma peak is observed, which shifts toward higher frequencies with increasing feedback gain (**e**) and V1 input gain (**f**). **g**,**h**, Coherence spectra: 3% contrast. A broad peak in the beta band is observed at low feedback gain, which vanishes at high feedback gain. No such peak is observed with changes in V1 input gain. **i**,**j**, Coherence spectra: 50% contrast. A beta peak is observed for low feedback gain, which shifts toward higher (gamma) frequencies with increasing feedback gain. No beta peak is observed for changes in V1 input gain, but the gamma peak shifts toward higher frequencies with increasing V1 input gain. **k**,**l**, Communication subspaces. Increasing feedback gain enhances inter-areal (circles) communication while decreasing within-area (squares) communication. Conversely, increasing V1 input gain decreases both inter- and within-area communication. The V2 input gain β_2 is held at its baseline value (1.0) throughout; the effect of varying β_2 is shown in Suppl. Fig. S8.

gain (right panels) is varied. These comparisons are presented for both a low (3%; Figs. 3c,d) and a high (50%; Figs. 3e,f) stimulus contrast. See Supplementary Material for comprehensive results across other contrast levels (Suppl. Fig. S1), and Suppl. Fig. S8d–i for the effect of modulating the V2 input gain.

The model predicts oscillatory activity in the alphafrequency range (≈ 10 Hz) for low stimulus contrasts and small feedback gain values (Fig. 3*c*). As the feedback gain to the principal cells increases, this alpha peak diminishes and ultimately vanishes at high feedback gain levels, while a broad peak concurrently emerges in the gamma frequency range (Figs. 3c,e). No such oscillatory activity in the alpha-frequency range is observed by manipulating the V1 input gain, for any contrast (Figs. 3d,f).

The power spectra are dominated by oscillatory activity in the gamma-frequency range at high stimulus contrast, regardless of the V1 input gain and feedback gain values (Figs. 3e,f). Increasing the feedback gain causes this gamma peak to shift towards higher frequencies. A similar shift of the gamma peak to higher frequencies is observed when the V1 input gain is increased. However, a key distinction arises in their impact on the peak's bandwidth: the gamma peak bandwidth narrows with increasing V1 input gain but broadens with increasing feedback gain.

Inter-areal coherence

Coherence measures the statistical correlation between two signals (e.g., LFPs in two different brain areas) as a function of frequency. Neuronal synchrony is hypothesized to contribute to the dynamic selection and routing of information both within and between brain areas [10, 38, 39].

Our model's predictions for V1-V2 coherence spectra accurately capture key experimental trends observed across various stimulus contrasts. We analytically determined the coherence between the maximally firing neurons of V1 and V2 (see Methods). Figure 2e displays the theoretically predicted V1-V2 coherence spectra for different contrasts. These model predictions are compared with experimental V1-V2 coherence data from macaque monkeys, obtained under identical contrast conditions (Fig. 2f; [31]). The model accurately captures three qualitative observations: 1) high coherence at lower frequencies for low contrasts; 2) a shift of the peak coherence toward higher frequencies with increasing contrast; and 3) a decrease in peak coherence after reaching a maximum around 40% contrast. The coherence peaks in the gamma frequency range are due to normalization; removing normalization eliminates the peaks in coherence (see Suppl. Fig. S4f). Suppl. Fig. S7b shows model fits to experimental measurements of the coherence spectra. Feedback and input gain differentially modulate V1-V2 coherence across contrasts (Figs. 3g-j; Suppl. Fig. S8j–o). At low contrast (3%), a coherence peak in the beta band (12-30 Hz) is observed with low feedback gain (Fig. 3g); this peak disappears as feedback gain is increased. Notably, at low feedback gain, a dip in V1-V2 coherence occurs in the alpha band, corresponding to the frequencies where peak power is observed in V1. This suggests that alpha band activity, despite its prominence in V1 power, may not effectively propagate to V2 or participate in inter-areal communication. No such peaks are observed with changes in the V1 input gain for low contrast (Fig. 3h). At high contrast (50%), increasing feedback gain shifts the peak coherence towards higher frequencies and increases both its magnitude and width (Fig. 3i). A similar frequency shift in peak coherence occurs with increasing V1 input gain, but its magnitude and width decrease (Fig. 3j). Increasing the V2 input gain shifts the coherence peak to higher frequencies, broadens it, and reduces its magnitude, and produces no beta-band peak at low contrast (Suppl. Fig. S8l,o).

Communication subspaces

Neural communication is hypothesized to be channeled through "communication subspaces (CS)" [15]. Although neural activity space is high-dimensional, neural communication may not utilize this vast space. CS are lower-dimensional manifolds within the broader neural activity space that preferentially engage in inter-areal communication. In contrast, within-area communication utilizes a distinct, higher-dimensional "private" subspace [15].

Our theory successfully captures key characteristics of the CS (Figs. 2g,h). We derived an analytical model to examine both inter-areal (V1-V2) and within-area (V1-V1) communication (see Methods). We found that the CS is low-dimensional, much lower than the number of sampled neurons (source and target populations of 30 neurons in Figs. 2g and 3k,l; see Methods); the dependence of the dimensionality on the number of modeled and sampled neurons is characterized in Suppl. Fig. S9. In addition, the inter-areal communication (V1-V2) subspace is lower-dimensional than the within-area (V1-V1) communication subspace. Figure 2g illustrates the model prediction performance (i.e., the ability to predict activity fluctuations of a target subpopulation from the activity fluctuations of a source subpopulation using a linear model) as a function of the dimensionality of the CS. These theoretical results are qualitatively consistent with the experimental data (Fig. 2h; [15]).

Feedback and input gain differentially modulate communication subspaces (Figs. 3k,l; Suppl. Fig. S8p– r). Increasing feedback gain from the higher cortical area (V2) to the lower cortical area (V1) enhanced inter-areal communication while diminishing within-area communication (Fig. 3k), without substantially altering the subspace dimensionality (see Suppl. Figs. S3a,b). Conversely, increasing V1 input gain reduced both the inter-area and within-area communication (Fig. 3l). These results are for 100% contrast; see Suppl. Fig. S3c for other contrasts. Increasing the V2 input gain, β_2, enhances inter-areal communication, like feedback gain, but leaves within-area communication approximately unchanged (Suppl. Fig. S8r). However, the effect of feedback gain is substantially larger than that of V2 input gain. The observed enhancement of inter-areal communication with increased feedback suggests that gain modulation, and top-down feedback in particular, can dynamically modulate the efficiency of information transfer between cortical areas without altering the underlying structure of the communication subspace (see below, Dynamic modulation of functional connectivity through feedback), a finding consistent with previous work on subspace dimensionality [40].

Frequency decomposition of communication subspaces

The theory predicts that the effectiveness of communication between brain areas varies systematically with frequency and contrast (Fig. 4). We quantified the efficacy of communication by analytically calculating the prediction performance for each frequency component (see Methods). Using the Wiener-Khinchin theorem, we converted the power spectra of the source and target population into their noise correlations. To isolate noise correlations at a specific frequency, we applied an idealized narrow-band pass filter. This conceptual filter is implemented via a transfer function, which is non-zero only within a very narrow frequency band around the frequency of interest. Applying this filter to the power spectra and integrating across the frequency domain yields the frequency-specific correlations (see Methods, “Frequency decomposition of the communication subspace”. This conversion enabled us to predict the power of the target population from the source population (using a linear readout) at each frequency, thus providing a measure of communication efficacy (prediction performance) as a function of frequency.

V1-V2 communication is dynamically modulated by stimulus contrast in a frequency-dependent manner. At low stimulus contrasts (e.g, approximately 10%), a distinct peak in communication is observed at ≈ 20 Hz (Fig. 4b). As stimulus contrast increased, this preferred frequency for communication shifts toward higher frequencies (Fig. 4b). Furthermore, the magnitude of this communication peak exhibits a non-monotonic relationship with contrast, reaching its maximum at around 40% contrast before declining at higher contrasts. These communication peaks are due to normalization; removing normalization eliminates the peaks in both coherence and prediction performance (see Suppl. Figs. S4f,g). Under conditions of very low contrast, no distinct peak in communication is observed; instead, communication is mostly concentrated at very low frequencies (< 10 Hz). These frequency-specific trends in communication mirror the V1-V2 coherence spectra (Fig. 4a). This correspondence is a direct consequence of how the two quantities are defined: coherence (Eq. 7) and frequency-resolved prediction performance (Eq.

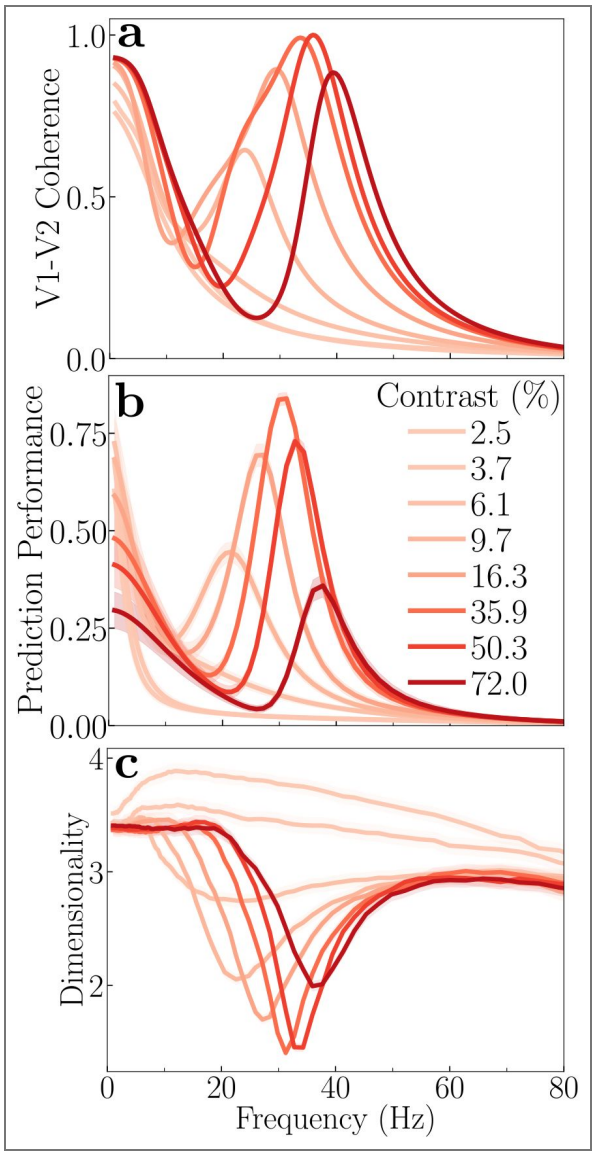


Figure 4. Theoretical prediction: Frequency-dependence of communication.

a, V1-V2 coherence spectra for different contrasts; same as in Fig. 2e. **b**, Prediction performance versus frequency at different contrasts, averaged across different subsets of the source and target populations. At low stimulus contrast (≈ 10 %), a peak in communication efficacy is observed around 20 Hz. As stimulus contrast increases, the preferred frequency of communication shifts towards higher frequencies (40 Hz). The magnitude of communication reaches a maximum around 40% contrast, before declining at higher contrast levels. Under conditions of very low contrast (< 10%), communication is mostly concentrated at very low frequencies. These frequency-specific trends in communication parallel the V1-V2 coherence patterns shown in panel **a. c**, Dimensionality of the communication subspace versus frequency, averaged across different subsets of the source and target populations. A notable dip in dimensionality occurs at frequencies corresponding to peaks in communication (panel **b**) and coherence (panel **a**). The shaded areas in panels **b** and **c** represent the standard error of the mean (SEM) across different subsets of the source and target populations.

22) are both computed from the same cross-spectral density matrix, so the frequencies at which the source and target populations are most strongly correlated are also those at which one best predicts the other.

The dimensionality of the CS is the number of dimensions of source activity required to reach the full-rank prediction performance (within a small tolerance); it is dictated by the rank of the covariance matrix between the source (V1) and target (V2) subpopulations (see Methods: Interpretation of the results, Suppl. Fig. S5). The dimensionality of the inter-areal (V1-V2) CS also varies with frequency and stimulus contrast (Fig. 4c). The dimensionality of the communication subspace decreases at frequencies where inter-areal coherence and communication are maximal, for each contrast (Fig. 4c).

The dimensionality of the CS is influenced mainly by three factors. First, dimensionality depends on normalization. Although dimensionality is relatively low in the absence of normalization (Suppl. Fig. S4h), including normalization further reduces dimensionality at coherent frequencies (Fig. 4c). The intuition for why normalization reduces dimensionality is that when only a few neurons are strongly driven, and their responses are highly correlated across areas (V1 and V2), normalization suppresses the activity of other nearby neurons that are not strongly driven. This suppressive interaction, akin to a soft winner-take-all mechanism, diminishes the contribution of other neurons to the shared covariance, thereby reducing the overall dimensionality. Second, the dimensionality is increased for a fully interconnected (all-to-all) connectivity between V1 and V2 (Suppl. Fig. S4l). Third, delocalized inputs (e.g., noise images or naturalistic images) lead to a higher dimensionality than localized inputs (e.g., gratings) (Suppl. Fig. S4p), commensurate with experimental results [15].

Dynamic modulation of functional connectivity through feedback

Anatomical connections influence neural communication but do not completely determine it. Rather, it has been observed experimentally that functional connectivity between brain areas changes dynamically [41, 42].

Our theory predicts that functional connectivity is controlled by modulating feedback strength (Fig. 5). We expanded our model to encompass three cortical areas: V1, V4, and V5→ (also called MT). In this expanded model, V1 has reciprocal connections with both V4 and V5, representing the parallel dorsal and ventral visual pathways, respectively [43]. For simplicity, we assume there are no direct connections between V4 and V5 in the model. Analysis of this three-area model provides a demonstration that differential feedback strength can control functional connectivity. When the feedback strength from V5 V1 is greater than the feedback from V4 → V1, we observe that V1 strengthens communication with V5 (Fig. 5a). Conversely, when the feedback from V4 to V1 is stronger, V1 shows enhanced communication with V4 (Fig. 5b). This suggests that the relative strength of feedback from the higher cortical area (V5 or V4) can dictate the information flow from the lower cortical area (V1). One possible way this could be achieved is by selectively enhancing the neural activity in either V5 or V4, which will, in turn, increase the feedback to V1. This is commensurate with experimental evidence that feature-based attention and task demands may selectively boost activity in one or another cortical area [44], and specifically that attending to motion increases the transfer of motion information (communication) between V1 and V5 [45]. Therefore, top-down signals can dynamically modulate functional connectivity from lower to higher cortical areas, simply by adjusting the gain of neural activity in one higher cortical area (e.g., V4) relative to another (e.g., V5). This mechanism enables flexible cortical information routing, which is essential for cognitive flexibility. Because increasing the V2 input gain also enhances inter-areal communication (Suppl. Fig. S8r), we expect that differential V2 input gain to V4 versus V5 would route communication in the same way although to a lesser extent; the two mechanisms are distinguished by their effect on within-area communication, which is reduced by feedback gain but unchanged by V2 input gain, and by their spectral signatures (Suppl. Fig. S8).

Note that, to isolate the variable of interest, we modeled V4 and V5 with identical parameters for their recurrent, feedforward (V1 → V4/V5) and feedback (V1 ← V4/V5) connectivity. This ensures that the observed increase in communication with either V4 or V5 is a direct consequence of the

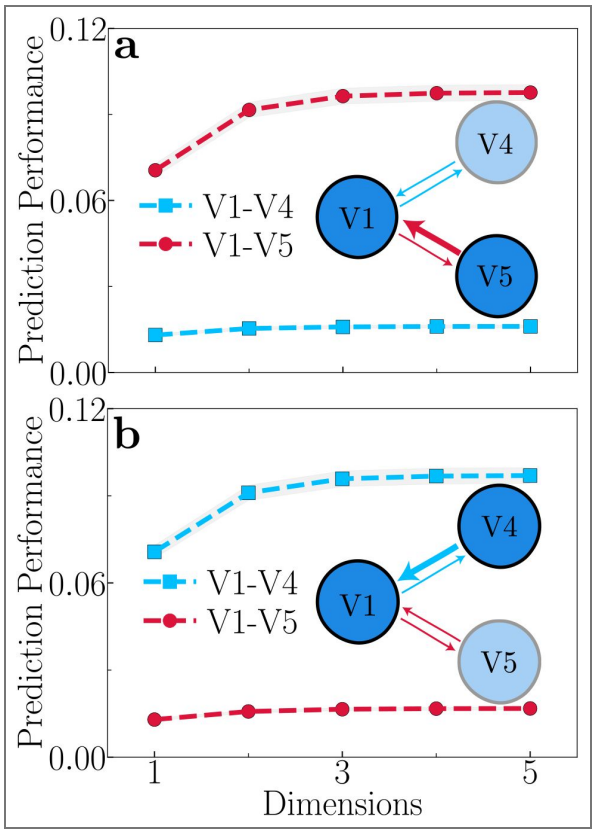


Figure 5. Theoretical prediction: Feedback-dependent modulation of functional connectivity.

a, Feedback from V5 → V1 is stronger than feedback from V4 → V1, and V1 preferentially enhances communication with V5. **b**, Feedback from V4 → V1 is stronger than feedback from V5 → V1, and V1 preferentially enhances communication with V4. The plotted prediction performance is an average across different subsets of the source and target populations, and the shaded areas represent the standard error of the mean (SEM). These results demonstrate that top-down feedback can dynamically route information flow between cortical areas, i.e., modulating functional connectivity.

stronger feedback from that specific area.

Discussion

We present an analytically tractable theory of neural circuit function that implements divisive normalization dynamically across interconnected brain areas. Emergent properties of the theory account for a diverse array of neurophysiological phenomena, including oscillatory activity (measurable as peaks in LFP power and coherence spectra) and communication subspaces. Based on the theory, we propose a unified view of inter-areal communication.

Our theory yields two key, experimentally-testable predictions. First, the dimensionality of the communication subspace is lowest at the frequencies at which V1-V2 coherence and communication peak, and these frequencies shift with contrast. Second, increasing feedback strength selectively enhances inter-areal communication and diminishes within-area communication, without altering the subspace dimensionality, whereas increasing the V2 input gain enhances inter-areal communication without reducing within-area communication. Candidate experimental manipulations and readouts for each of these predictions, and for the gain manipulations discussed below, are given in "Experimental predictions of the model". A comprehensive list of further model results and their robustness is provided in *SI Appendix* section "Robustness with respect to model parameters".

Furthermore, the theory proposes a mechanism for dynamic functional connectivity: top-down feedback signals enhance communication between brain areas simply by adjusting the gain of neural activity in one or more cortical areas. This modulation of neural gain offers a mechanism for dynamically changing the functional connections between brain areas over time.

Neural circuit function

The original ORGaNICs model [6, 7], in its initial form without feedback connections, serves as a biophysically-plausible counterpart to Long Short Term Memory units (LSTMs) [46], a class of artificial recurrent neural networks (RNNs) prevalent in machine-learning applications [47]. Unlike LSTMs or any other previously proposed RNNs, ORGaN-ICs have built-in recurrent normalization, which offers several advantages, including stability even with exuberant recurrent connections [8, 9], and long timescales without fine-tuning [24].

The hierarchical version of the ORGaNICs theory introduced here adds feedback connections, drawing a partial analogy to "hierarchical predictive coding" models of sensory and perceptual processing [48–53]. However, our hierarchical theory diverges from predictive coding in its fundamental mechanisms. In predictive coding, it is proposed that forward-propagating signals represent prediction errors, while feedback pathways convey predictions to lower levels of hierarchy. The goal is to minimize prediction error by matching sensory input with top-down predictions. In hierarchical ORGaNICs, instead of propagating prediction errors, the circuit balances the input, recurrent and feedback drives of each principal cell (Eq. 1a) so that the steady-state responses follow the normalization equation (Eq. 2). This mechanism is more consistent with the observed physiology than predictive coding models. For example, neurons maintain sustained activity in response to a constant, fully predictable stimulus [54], whereas in predictive coding the forward-propagated activity, representing prediction error, should be "explained away" as the stimulus becomes predictable.

Normalization

The proposed theory admits an analytical fixed point that follows the normalization equation (Eq. 2) *exactly*, for arbitrary non-negative normalization weights, when the input drive is constant over time, the recurrent weight matrix is the identity, and the input gain and feedback gain are equal to one ($\beta = \gamma = 1$) (see *SI Appendix* for derivation) [8, 9]. For the center-surround recurrent weights used here, the fixed point approximates normalization closely (Methods). Several circuit mechanisms have been hypothesized to underlie normalization, including shunting inhibition, synaptic depression, and inhibition-stabilized networks [26, 55, 56]In these models, normalization

is realized either largely feedforward, through a divisive conductance or synaptic depression at the input, or approximately, as an emergent property of recurrent excitatory–inhibitory dynamics. The present circuit differs in three respects: the normalization equation, with an arbitrary normalization pool, is a fixed point of the recurrent dynamics rather than a description fitted to their steady state; the dynamics remain stable even with exuberant recurrent connections [8, 9]; and, because the fixed point and the linearized dynamics about it are available, the same circuit yields power spectra, coherence, and communication subspaces in closed form (Methods), which is what allows the spectral and population-level phenomena examined here to be traced to a single mechanism.

Our model achieves normalization dynamically via recurrent amplification, which is inversely regulated by the activity of inhibitory modulator cells. When the input is weak, the modulator cells exhibit a small response, leading to strong recurrent amplification. Conversely, a strong input generates a large response from modulator cells, which suppresses the amplification. This process modulates both excitatory and inhibitory recurrent signals, which is consistent with the experimental observation that normalization involves a decrease in both recurrent excitatory and inhibitory conductances [26]. Because the theory implements normalization, it is commensurate with a wide range of physiological and psychophysical phenomena previously explained by the normalization model (see Introduction for references).

The theory links normalization with oscillatory activity. Removing divisive normalization eliminates oscillatory peaks in power and coherence spectra (Suppl. Figs. S4e,f) and abolishes the link between coherence, prediction performance, and subspace dimensionality (Suppl. Figs. S4g,h), demonstrating that normalization is central to both the dynamics and function of the circuit.

Role of feedback

This paper provides an experimentally testable computational model of cortical function where neural activity is shaped by a combination of feedforward drive (bottom-up) and feedback drive (top-down). Feedback in the model may serve a variety of functions. First, feedback provides a means for controlling inter-areal communication (functional connectiv-ity) simply by changing the gain of neural activity in one or more cortical areas (see Discussion section “Attention and functional connectivity”). Second, the feedback drive may contribute to selective attention (see Section “Attention and Functional Connectivity”). Third, the relative strength of the input gain and feedback gain determines whether neural responses are driven bottom-up, top-down, or a combination of the two. Fourth, feedback contributes to the local normalization signal by providing excitatory drive to inhibitory neurons (see Methods). This feedback-mediated inhibition mechanism is essential for maintaining network stability, and is commensurate with experimental evidence that normalization depends in part on loops through higher visual cortical areas [57].

The interplay between the input gain and the feedback gain determines the primary driver of neural responses. If the input gain is strong and the feedback is weak, the system behaves like traditional feedforward models. Conversely, with strong feedback and weak input, the network is hypothesized to behave like generative models, creating sensory representations from higher-level abstract knowledge, akin to recalling an image from memory. When input and feedback are balanced, the framework may integrate prior beliefs with sensory evidence (see [5] for details), akin to models based on Bayesian inference [58, 59].

The theory predicts differential effects of changing input gain and feedback gain on the power, coherence spectra, and inter-areal communication (see Fig. 3 and Suppl. Fig. S8). Hence, these metrics may be used as a signature for distinguishing changes in input gain from changes in feedback (“Experimental predictions of the model”).

Unified view of inter-areal communication

Our theoretical framework offers a novel perspective on inter-areal communication. Rather than viewing “communication through coherence” (CTC) and “communication subspace” (CS) as alternative hypotheses, our theory unifies them, showing that coherence and communication

subspaces are complementary empirical phenomena emerging naturally from the same underlying mechanism. A similar conclusion was recently reached by Ni et al. [60] using a large-scale spiking neural circuit model. The frequency decomposition of the communication subspace mirrors the inter-areal coherence at every contrast. Communication is strongest at the frequencies of peak V1-V2 coherence, and the communication subspace is lowest-dimensional at the same frequencies. The reason is that coherence, prediction performance, and dimensionality are all computed from the same cross-spectral density matrix. Normalization shapes this matrix in two ways. First, normalization creates a resonance. This resonance sets the peaks of the coherence and of the prediction performance. Second, normalization suppresses weakly driven neurons. This suppression lowers the rank of the cross-spectral density matrix at the resonant frequencies and thereby reduces the dimensionality of the communication subspace.

Attention and functional connectivity

The theory provides a possible explanation for the efficient routing of sensory information between cortical areas, i.e., for changing functional connectivity. Lower cortical areas, such as V1, can flexibly relay information to specific higher cortical areas (e.g., V4 vs. V5) based on the relative strength of the feedback from each of the higher cortical areas. The strength of the feedback can be controlled either by changing the feedback gain or (perhaps more simply) by changing the gain of the neural responses in each of the higher cortical areas. We hypothesize that this feedback-mediated process accounts for changes in functional connectivity [61] and feature-based attention [62]. Consistent with this hypothesis, experimental evidence demonstrates that feature-based attention is spatially global, i.e., that neural activity is boosted in feature-selective neurons with receptive fields covering the visual field [63].

The theory may also provide insights into the neural processes underlying spatially-selective attention with the addition of a retinotopic map. The theory identifies two distinct pathways through which attention modulates neural gain. First, attention may modulate the input gain of the sensory neurons, consistent with the established normalization model of attention [64]. Second, attention may modulate the activity of the sensory neurons via the top-down feedback drive. Exogenous attention is the fast, involuntary orienting of attention driven by a salient stimulus, and endogenous attention is the slower, voluntary, goal-directed deployment of attention. We hypothesize that exogenous attention acts primarily through the input gain and endogenous attention primarily through the feedback drive. This mapping is a testable hypothesis since the two pathways produce divergent spectral signatures (“Experimental predictions of the model”).

Thalamo-cortical model

Alpha oscillations are a defining characteristic of thalamo-cortical dynamics, particularly prominent within the visual system during states of “relaxed wakefulness” [65, 66]. Specifically, alpha oscillations are hypothesized to be generated and modulated by thalamo-cortical loops [66, 67], suggesting that alpha rhythms signify a state of functional inhibition or “idling” in sensory pathways, potentially gated by corticothalamic feedback [67, 68]. While the hierarchical model presented in this paper was designed for cortico-cortico interactions, it may be adapted to model the thalamo-cortical interactions that contribute to the generation of alpha waves. Our theory predicts the emergence of oscillatory activity in the alpha band (8-12 Hz) under conditions of low stimulus contrast, low feedback to principal cells, and high feedback to inhibitory modulator cells ($\gamma < 1$, Figs. 3c,g). This aligns well with the existing literature demonstrating the role of alpha oscillations in mediating cortical inhibition [68, 69]. It is, however, unclear if the current theory will suffice, on its own, to fully capture the phenomenology of alpha waves. Furthermore, we hypothesize that the input gain modulator responses depend in part on thalamocortical loops, hence positioning the theory as a means for modeling the interplay between bottom-up sensory input and top-down feedback within the thalamocortical circuit, the gating of visual information flow, thalamic control of functional cortical connectivity [70] and thalamocortical loops in other neural systems [71].

Limitations

Our model assumes instantaneous feedback, whereas biological feedback is delayed due to axonal conduction, synaptic transmission, and processing in higher cortical areas [72, 73]. Examining how inter-areal synchronization depends on feedback delay is an important direction for future work.

The feedforward, feedback, and recurrent weight matrices may contain both positive and negative weights. We therefore posit inhibitory interneurons to implement the sign inversion, corresponding to negative synaptic weights. These inhibitory neurons need not be one-to-one with their excitatory inputs. Rather, each may compute a partial weighted sum of their inputs corresponding to the negative terms in the matrix multiplications. Regardless, these inhibitory interneurons will introduce an additional delay. The weight matrices are thus effective (net) connectivity, and the principal cells, which are excitatory in the hypothesized mapping onto pyramidal neurons ("Biological plausibility"), interact through mixed-sign effective weights both recurrently and across areas. The circuit as written therefore does not respect Dale's law at the level of individual connections, but it could be elaborated to do so.

Although most computations in the theory are biophysically plausible (see "Biological plausibility" and [7]), some mechanisms remain unresolved. In particular, we can only speculate about the possible cellular mechanisms for the element-wise products in Eqs. 1b and 1c (see "Biological plausibility").

Thetime constants used in simulations are unreal-istically short for individual neurons (see Table S1). Using more realistic values would shift gamma-band oscillations to lower frequencies. When all time constants were released as free parameters to fit to the experimental measurements of the power and coherence spectra (Suppl. Fig. S7), they converged to values between 0.2 and 3 ms. Multiplying all the time constants (those of the four cell types and of their firing rates, those of the noise processes, and also $\alpha_{1,2}$) by a common factor k would decrease the frequencies of the spectral peaks by that same factor k. The fixed point would be unchanged, every eigenvalue of the Jacobian would be divided by k, and the shape of every spectral feature would be preserved. For example, with τ = 10 ms the spectral peaks of Fig. 2c,e would lie at 4–6 Hz, but the rightward shift with contrast, the non-monotonic peak amplitude, the $1/f^4$ fall-off and all gain-dependent trends would be preserved. The time constants may therefore be better interpreted as effective population time constants rather than as the intrinsic time constants (capacitance/conductance) of individual neurons.

Finally, the model omits neural adaptation, a ubiquitous feature of cortical responses. Incorporating adaptation into the hierarchical framework remains an important direction for future studies.

Biological plausibility

We hypothesize a correspondence between the variables in the theory and cortical microcircuits composed of identified cell types. Principal cells **y** are hypothesized to correspond to pyramidal neurons; inhibitory modulatory cells **a** to a combination of parvalbumin-expressing (PV+) and somatostatin-expressing (SST+) interneurons; and inhibitory interneurons **q** to vasoactive intestinal polypeptide (VIP) neurons.

In the proposed circuit (Fig. 1), Principal cells **y** excite both inhibitory modulatory cells **a** (via **u**) and inhibitory interneurons **q**. VIP-like cells **q** inhibit modulatory cells **a**, which in turn inhibit principal cells **y**. This architecture closely mirrors known cortical motifs: pyramidal neurons excite PV+, SST+, and VIP cells; VIP cells inhibit SST+ cells; and both PV+ and SST+ cells inhibit pyramidal neurons [75–79]. We further hypothesize that modulator cells **a** have large receptive fields and broad orientation-selectivity (reflecting properties of the normalization pool), consistent with the response properties of SST+ and PV+ neurons, respectively [76, 80, 81].

We also posit excitatory feedback from higher cortical areas to both principal cells **y** and modulatory cells **a** (Fig. 1). Feedback to principal cells is likely received by the distal apical dendrites of layer 1, while feedback to modulator cells **a** is commensurate with experimental

evidence that feedback from higher cortical areas targets SST+ neurons [82–84], and that normalization depends in part on loops through higher visual cortical areas [57].

One potential discrepancy between the model and the cortical circuitry is that excitatory input from principal cells **y** to modulatory cells **a** is mediated by an intermediate excitatory population **u**, whereas pyramidal neurons directly excite PV+ and SST+ cells in cortex. However, the theory can be reformulated so that principal cells provide direct excitatory input to modulatory cells by substituting Eq. 1b into Eq. 1c (see Methods), without altering the model's functional behavior.

We hypothesize that the computations performed by principal cells correspond to processes distributed across dendritic compartments of pyramidal neurons (Fig. 6). The equation governing the responses of the principal cells (see Fig. 6 and Methods, Eq. 1a) expresses the following synaptic current: (input gain) (input drive) + (recurrent gain) [recurrent drive + (feedback gain) (feedback drive)]. The input drive is computed in the distal basal dendrites, commensurate with evidence that feedforward synaptic inputs contact basal dendrites. The input gain could be controlled by inhibitory synapses near the basal trunk. The feedback drive is computed in distal apical dendrites in layer 1, which are known to receive signals from higher cortical areas. The feedback gain might be attributed to the amplification of the feedback drive within the apical trunk of pyramidal cells [85]. The recurrent drive is likely computed at more proximal apical dendrites, with recurrent gain shaped by inhibitory inputs near the proximal apical trunk.

The weighted sums and gain control may be implemented via balanced excitatory–inhibitory push–pull and push–push mechanisms, respectively [86, 87]. Alternatively, various other mechanisms for gain control have been identified [88]. We can, however, only speculate about the possible cellular mechanisms for the element-wise products in Eqs. 1b and 1c. The element-wise product between presynaptic firing rates in Eq. 1b might be mediated by coincident spike detection, e.g., assuming that each of the spike trains is Poisson and that the synaptic current is proportional to the probability of coincident spikes within a short time window [89]. The element-wise product between postsynaptic membrane potential and presynaptic firing rate in Eq. 1c might be mediated by NMDA receptors, given that the synaptic current of an NMDA channel is jointly dependent on the presynaptic firing rate and the local membrane depolarization, in an approximately multiplicative manner [90, 91].

Experimental predictions of the model

The theory makes several predictions that can be tested with simultaneous recordings from two or more cortical areas. For each, we outline a candidate experimental manipulation and the readouts expected. The manipulations are speculative and are offered as starting points rather than established procedures.

Frequency-resolved communication subspace (Fig. 4)

The dimensionality of the V1-V2 communication subspace is predicted to be lowest, and the prediction performance highest, in the frequency band of peak V1-V2 coherence, and this band is predicted to shift to higher frequencies with stimulus contrast. The test requires simultaneous V1 and V2 population recordings across a range of contrasts. From the cross-spectral density matrix of the recorded populations, the communication subspace, prediction performance and dimensionality can be computed at each frequency by reduced-rank regression, without band-pass filtering (Methods, "Application to experimental measurements").

Routing by differential feedback (Fig. 5)

V1 is predicted to communicate preferentially with whichever higher area provides the stronger feedback. With simultaneous recordings from V1, V4 and V5, the relative feedback strength could be varied behaviorally, by having the subject attend to motion versus form [44, 63, 92, 93], or directly, by activating or inactivating V4 or V5 (e.g., optogenetically, electrically or pharmacologically).

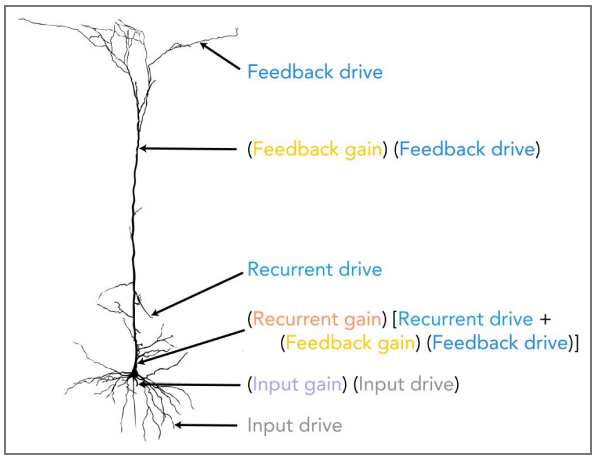


Figure 6. Hypothesized mapping of computational components in the model onto the dendritic compartments of a pyramidal cell.

The input drive is hypothesized to be the weighted sum of feedforward inputs arriving at the distal basal dendrites. This is modulated by an input gain, corresponding to inhibition at the proximal basal dendritic trunk. In the apical tuft, the feedback drive represents a weighted sum of feedback inputs from higher cortical areas. This signal is amplified by a feedback gain within the distal apical trunk. The recurrent drive corresponds to the sum of lateral inputs on the proximal apical dendrites. The combined recurrent and feedback signals are then modulated by a recurrent gain at the proximal apical trunk. The overall synaptic current is a combination of these gain-modulated drives. The pyramidal neuron cell shown is from [74].

Normalization (Suppl. Figs. S4 and S5)

We hypothesize that the inhibitory modulator cells **a** correspond to PV+ and SST+ interneurons ("Biological plausibility"). Manipulating these interneurons so as to weaken normalization should reduce the gamma peaks in the power and coherence spectra, remove the frequency dependence of prediction performance, and increase the dimensionality of the communication subspace.

V1 input gain β_1 (Fig. 3, right column; Suppl. Fig. S8, middle column)

We hypothesize that the input gain is controlled by inhibitory synapses near the basal dendritic trunk of pyramidal neurons ("Biological plausibility"); identifying the interneuron type that provides this inhibition, and manipulating it, would be one way to change β_1 independently of the stimulus. Increasing β_1 should shift the contrast-response functions of both V1 and V2 along the log-contrast axis, shift the gamma peak to higher frequencies and narrow it, produce no alpha peak at low contrast, and decrease both inter- and within-area communication, leaving the subspace dimensionality unchanged.

Feedback gain γ_1 (Fig. 3, left column; Suppl. Fig. S8, left column)

Feedback gain could be manipulated by activating or inactivating (e.g., by cooling or optogenetic silencing) the higher area that provides the feedback (V2 or V4), or, more speculatively, by targeting the amplification of the feedback drive in the apical trunk of pyramidal neurons ("Biological plausibility"). Increasing γ_1 should produce a larger change in contrast gain in V2 than in V1; at low contrast, the alpha peak in V1 power and the beta peak in V1-V2 coherence that are present at low feedback gain should vanish; the gamma peak should shift to higher frequencies and broaden; and inter-areal communication should increase while within-area communication decreases, with the subspace dimensionality unchanged.

V2 input gain β_2 (Suppl. Fig. S8, right column)

The V2 input gain scales the V1 → V2 drive. It could be manipulated by targeting the V1 → V2 projection (e.g., optogenetic modulation of V1 axon terminals in V2) or the corresponding basal-trunk inhibition of V2 pyramidal neurons, with V1 and V2 recorded simultaneously. Its signature differs from those of the other two gains: V1 responses are unchanged while V2 responses are enhanced; the gamma peak shifts to higher frequencies, broadens, and decreases in amplitude; there are no alpha or beta peaks at low contrast; and inter-areal communication increases while within-area communication is unchanged.

Methods

The Model

We propose a multistage hierarchical recurrent circuit that incorporates feedforward connections from the primary visual cortex (V1) to higher visual areas (e.g., V2, V4, MT) and feedback connections from these higher brain areas to V1 (Fig. 1). We refer to V2 throughout this paper for concreteness, but it is merely an example of a higher brain area that is reciprocally connected with V1. This model builds on the ORGaNICs framework developed by Heeger et al. [6, 7].

The response of principal neurons in the model is a sum of three key components: i) input drive, a weighted sum of neural responses from a brain area that provides feedforward input (e.g., LGN → V1 and V1 → V2), modulated by input gain; ii) feedback drive, a weighted sum of neural responses from the higher cortical area (e.g., V1 ← V2 or V1 ← V4), modulated (viz., scaled) by feedback gain; and iii) the sum of recurrent drive, a weighted sum of neuronal responses from other neurons in the same brain area (e.g., V1 → V1 and V2 → V2) and scaled feedback drive, collectively modulated by recurrent gain.

The proposed neural circuit comprises four distinct cell types: the principal neurons (**y**), the modulatory excitatory neurons (**u**), the modulatory inhibitory neurons (**a** and **q**), with the following dynamics:

$$\tau_{y_1}\frac{\mathrm{d}\mathbf{y}_1}{\mathrm{d}t} = -\mathbf{y}_1 + \frac{\beta_1}{2}\odot\mathbf{z}_1 + \left(\frac{1}{1+\mathbf{a}_1^+}\right)\left(\mathbf{W}_{11}\sqrt{\mathbf{y}_1^+} + \frac{\gamma_1}{2}\odot\mathbf{W}_{12}\sqrt{\mathbf{y}_2^+}\right) \tag{1a}$$

$$\tau_{u_1}\frac{\mathrm{d}\mathbf{u}_1}{\mathrm{d}t} = -\mathbf{u}_1 + \left(\frac{\sigma_1}{2}\right)^2 + \mathbf{N}_1\left(\mathbf{y}_1^+\odot\mathbf{u}_1^{+2}\right) \tag{1b}$$

$$\tau_{a_1}\frac{\mathrm{d}\mathbf{a}_1}{\mathrm{d}t} = -\mathbf{a}_1 + \frac{\mathbf{W}_{12}\sqrt{\mathbf{y}_2^+}}{2\cdot\mathbf{q}_1^+} + \mathbf{u}_1^+ + \mathbf{a}_1\odot\mathbf{u}_1^+ + \alpha_1\frac{\mathrm{d}\mathbf{u}_1}{\mathrm{d}t} \tag{1c}$$

$$\tau_{q_1}\frac{\mathrm{d}\mathbf{q}_1}{\mathrm{d}t} = -\mathbf{q}_1 + \sqrt{\mathbf{y}_1^+}. \tag{1d}$$

The above equations are for V1; V2 follows a similar set of equations (see *SI Appendix*). The membrane potentials of neurons in cortical areas V1 and V2 are denoted by $\mathbf{y}_1$ and $\mathbf{y}_2$, respectively, while their corresponding firing rates are represented by $\mathbf{y}_1^+$ and $\mathbf{y}_2^+$. The $\odot$ symbol represents the Hadamard product, element-wise multiplication of vectors. To estimate the firing rates $\mathbf{y}_1^+$ and $\mathbf{y}_2^+$ from their respective membrane potentials, we employ a modified Gaussian Rectification (GR) model [94, 95]. The variable $\mathbf{z}_1$ denotes the input drive to V1 from the preceding brain area (the lateral geniculate nucleus of the thalamus, LGN), computed as a weighted ($\mathbf{W}_{zx}$) sum of the LGN firing rates (Fig. 7a illustrates the resulting orientation tuning curves). $\tau_{\mathbf{k}}$ is the time constant for each subtype of neuron, where $\mathbf{k} = \mathbf{y}, \mathbf{u}, \mathbf{a}, \mathbf{q}\cdot\mathbf{W}_{11}\sqrt{\mathbf{y}_1^+}$ and $\mathbf{W}_{12}\sqrt{\mathbf{y}_2^+}$ are the recurrent and feedback drive, respectively, where $\mathbf{W}_{11}$ is the recurrent weight matrix for V1. $\mathbf{W}_{11}$ has a center-surround structure where the closer connections are excitatory and the more distant connections are inhibitory (Fig. 7b), same as in [7]. The recurrent weight matrix for V2, $\mathbf{W}_{22}$, has a similar structure. The maximum eigenvalue for these recurrent weight matrices is set to unity to ensure stability. However, we have shown that this need not be the case because normalization ensures stability [8, 9]. $\mathbf{W}_{12}$ represents the inter-areal feedback connectivity matrix from V2 to V1; it is diagonally dominant, with net excitatory entries near the diagonal and weak net inhibitory entries farther from it (Fig. 7c). The feedforward matrix from V1 to V2, $\mathbf{W}_{21}$, is taken to be the transpose of $\mathbf{W}_{12}$. Given that $\mathbf{W}_{12}$ is symmetric, it follows that $\mathbf{W}_{12} = \mathbf{W}_{21}$. The normalization matrix in V1 ($\mathbf{N}_1$) and V2 ($\mathbf{N}_2$) is taken to be all ones, meaning each neuron contributes equally to the normalization. All of these matrices describe effective connectivity; negative weights may, for example, be implemented with disynaptic inhibition through interneurons that are not modeled explicitly (see Discussion). The input gain and feedback gain for V1 are modulated by β_1 and γ_1, respectively. The recurrent and feedback amplifications in V1 are modulated dynamically by the inhibitory neuron population $\mathbf{a}_1$.

Both the V1 and V2 neurons follow the normalization equation *exactly* when β and γ are set to unity (i.e. when there is identical feedback to principal neurons ($\mathbf{y}_1$) and inhibitory modulator ($\mathbf{a}_1$) neurons) and $\mathbf{W}_{11}$ is the identity matrix (i.e., each neuron recurrently amplifies itself), but the circuit closely approximates normalization for a wide range of parameters. For V1 neurons, normalization is expressed as:

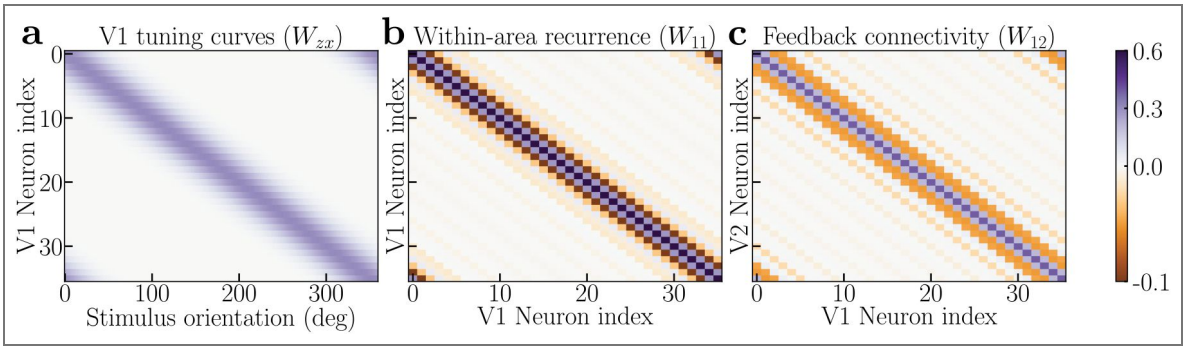


Figure 7. Connectivity matrices in the hierarchical recurrent circuit model.

a, Orientation tuning curves corresponding to the V1 encoding matrix ($\mathbf{W}_{zx}$). Each row represents the tuning curve of a V1 neuron. The weight matrix itself has positive and negative synaptic weights like standard models of orientation-selective receptive fields. **b**, The recurrent connectivity matrix for V1 ($\mathbf{W}_{11}$). The V2 recurrent connectivity matrix $\mathbf{W}_{22}$ has a similar structure. **c**, The Feedback connectivity matrix from V2 to V1 ($\mathbf{W}_{12}$). The corresponding feedforward matrix from V1 to V2, ($\mathbf{W}_{21}$), is its transpose. Since $\mathbf{W}_{21}$ is symmetric, the feedforward and feedback matrices are identical ($\mathbf{W}_{12} = \mathbf{W}_{21}$).

$$\mathbf{y}_1^+ = \frac{\lfloor \mathbf{z}_1 \rfloor^2}{\sigma^2 + \mathbf{N}_1 \lfloor \mathbf{z}_1 \rfloor^2}, \tag{2}$$

and likewise for V2

$$\mathbf{y}_2^+ = \frac{\lfloor \mathbf{z}_2 \rfloor^2}{\sigma^2 + \mathbf{N}_2 \lfloor \mathbf{z}_2 \rfloor^2}, \tag{3}$$

where $\mathbf{z}_2 = \mathbf{W}_{21}\mathbf{y}_1^+$ is the drive from V1 to V2, σ is the semi-saturation constant. Normalization is achieved dynamically via recurrent amplification (amplifying weak inputs but not strong inputs) with the three modulatory neurons, $\mathbf{u}_1$, $\mathbf{a}_1$ and $\mathbf{q}_1$ whose dynamics are governed by Eq. 1b, Eq. 1c, and Eq. 1d respectively. The equations for modulatory neurons are derived by substituting the normalization equation (Eq. 2) in the equation for the principal neurons (Eq. 1a) (see *SI Appendix* for derivation). An increase in the activity of principal neurons ($\mathbf{y}_1$) leads to enhanced activation in $\mathbf{u}_1$ and $\mathbf{q}_1$. This, in turn, drives a net increase of activity in modulatory neurons ($\mathbf{a}_1$). The elevated $\mathbf{a}_1$ activity subsequently exerts an inhibitory effect on $\mathbf{y}_1$, thereby implementing normalization by the recurrent interactions between the neural populations (see Appendix for derivation).

The parameter γ_1 modulates the feedback gain to the principal neuron population ($\mathbf{y}_1$). This determines the relative feedback to the principal neurons ($\mathbf{y}_1$) compared to the inhibitory modulator neurons ($\mathbf{a}_1$). As a rule of thumb, $\gamma_1 < 1$ indicates greater feedback inhibition, $\gamma_1 = 1$ denotes a balance between feedback excitation and inhibition, and $\gamma_1 > 1$ signifies more feedback excitation. For simplicity, we have set the feedback to the inhibitory neurons ($\mathbf{a}_1$) at a fixed value of 0.50 (hence, the factor of $\frac{1}{2}$ in equation Eq. 1c and Eq. 1a, see *SI Appendix*). Similarly, the parameter β_1 modulates the relative input gain to the principal neuron population ($\mathbf{y}_1$) compared to the modulatory excitatory neuron population ($\mathbf{u}_1$). For simplicity, we set the input gain to the modulatory neurons ($\mathbf{u}_1$) to 0.50 (hence, the factor of $\frac{1}{2}$ in equation Eq. 1a and Eq. 1b). Note that when β_1 and γ_1 are both set to unity and the recurrent matrix is identity, the network achieves excitatory-inhibitory balance exactly, and the normalization (Eq. 2) is followed precisely.

We modeled each brain area using a ring model, where principal neurons are arranged in a circle to represent continuous features like stimulus orientation. The recurrent connectivity between these neurons is determined by the distance between them (Fig. 7b). Each modeled area consists of 72 neurons for each of the four specified types ($\mathbf{y}_1$,$\mathbf{u}_1$,$\mathbf{a}_1$,$\mathbf{q}_1$). To simulate sensory input, a visual stimulus with a defined contrast and orientation is presented. The input drive is the product of the stimulus vector with the encoding matrix, whose rows are the tuning curves of the V1 principal neurons (Fig. 7a). The tuning curves are modeled as raised cosine functions, each peaking at the neuron's preferred orientation.

Numerical computation of fixed points

For all results reported here, the fixed point was computed numerically by integrating the deterministic dynamics to steady state. Starting from small uniform initial conditions, the dynamics were integrated with an adaptive Runge–Kutta scheme (RK45; relative and absolute tolerances of 10^{-10}) over a window spanning several thousand times the slowest time constant. The steady state was accepted when every state variable changed by less than 10^{-6} between consecutive output samples, and the Jacobian at the fixed point was verified to have eigenvalues with negative real parts (i.e., a stable fixed point).

The numerically computed fixed points closely approximate Eq. 2, to within 0.6% with the center-surround recurrent weight matrix (not the identity) and the feedback gain γ_1 set to one. The deviation from the normalization fixed point increases as the feedback gain γ_1 and the input gain β_1 deviate from 1.0.

Power and coherence spectra

The time evolution of neural population activity in V1 and V2 can be described by a general stochastic differential equation (SDE) with an additive noise term $\mathbf{L}d\mathbf{W}$:

$$d\mathbf{x} = \mathbf{f}(\mathbf{x}, \mathbf{u})dt + \mathbf{L}d\mathbf{W}, \tag{4}$$

where $\mathbf{x} \in \mathbb{R}^n$ is a vector representing the activity of neurons; $\mathbf{f(x, u)} \in \mathbb{R}^n$ represents the deterministic part of the dynamics explained above (Eq. 1); $d\mathbf{W}$ is a vector of independent Gaussian increments with correlation matrix $\mathbf{D}$ ($E[d\mathbf{W} \cdot d\mathbf{W}^T]$), and $\mathbf{L} \in \mathbb{R}^{n \times n}$ is the dispersion matrix defining how noise enters the system. Given the fixed point of the dynamical system, we can linearize the system around the fixed point ($\mathbf{x}^*$):

$$d\mathbf{x} = \mathbf{J}_{\mathbf{x}^*}(\mathbf{x} - \mathbf{x}^*)\,dt + \mathbf{L}d\mathbf{W}, \tag{5}$$

where $\mathbf{J}_{\mathbf{x}}*$ is the Jacobian computed at the fixed point $\mathbf{x}^*$. This linearization leads to a wide-sense stationary Gaussian process, $\mathbf{x}(t)$, whose power spectral density matrix, $\boldsymbol{S}(\omega) \in \mathbb{R}^{n \times n}$, is given by [96]:

$$\mathcal{S}(\omega) = \mathbb{E}\left[|\hat{\mathbf{x}}(\omega)|^2\right] = (\imath\omega\mathbf{I} + \mathbf{J})^{-1}\mathbf{L}\mathbf{D}\mathbf{L}^\top(-\imath\omega\mathbf{I} + \mathbf{J})^{-\top} \tag{6}$$

From $\boldsymbol{S}(\omega)$, we can directly compute the coherence, κ_{ij}, between any two i and j neurons in the two areas:

$$\kappa_{ij} = \frac{|\mathcal{S}_{ij}|^2}{\mathcal{S}_{ii}\mathcal{S}_{jj}}. \tag{7}$$

The linearization was validated by direct numerical integration of the full nonlinear stochastic system, including all noise sources, with the Euler–Maruyama scheme (time step 10 μs; 100 independent trials of 16 s after a 1 s burn-in, at each contrast); power and cross-spectra were estimated from the simulated trajectories by Welch's method (4 s segments, 50% overlap) and compared with Eqs. 6 and 7 (Suppl. Fig. S6; *SI Appendix*, "Numerical validation of the analytical spectra"). A quantitative comparison with experimental spectra, in which the model parameters are fitted, is described in the *SI Appendix* ("Quantitative fit to experimental power and coherence spectra") and shown in Suppl. Fig. S7.

Communication subspaces

Here we present the method used to characterize the communication subspaces between two subpopulations (source and target) of neurons [15]. These subpopulations may comprise distinct neurons in the same brain area or across different brain areas. The basic idea is to determine how well the fluctuations in responses of a target subpopulation can be predicted using a linear transform (a weighted sum) of the fluctuations in responses of the source subpopulation, and to determine the dimensionality of this linear transform.

The fluctuations in the responses and their co-variability are fully determined by the correlation matrix at steady-state when the strength of the noise is small. Therefore the correlation matrix, denoted as $\mathbf{C}(0)$, for the responses $\mathbf{x}(t)$ can be computed by solving the Lyapunov equation [96]:

$$\mathbf{J}\mathbf{C}(0) + \mathbf{C}(0)\mathbf{J}^\top = -\mathbf{L}\mathbf{D}\mathbf{L}^\top. \tag{8}$$

Once $\mathbf{C}(0)$ is obtained, we follow the method proposed by Semedo et al. [15] to construct an analytical model of the communication subspace. To achieve this, we divide the mean-

subtracted responses of the entire neuronal population, denoted as $\mathbf{x}$, into two subsets: $\mathbf{s}$ (responses of the source neurons, e.g., in V1) and $\mathbf{t}$ (responses of the target neurons, e.g., in V2 or V1). We consider a linear model for predicting target responses using the source responses, $\hat{\mathbf{t}} = \mathbf{B}^\top \mathbf{s}$, where the mean squared error (MSE) is given by:

$$\varepsilon = \mathbb{E}\left[\left|\mathbf{t} - \mathbf{B}^\top \mathbf{s}\right|^2\right] = \mathrm{Tr}\left(\mathbf{C}_2 + \mathbf{B}^\top \mathbf{C}_1 \mathbf{B} - 2\mathbf{B}^\top \mathbf{C}_3\right), \tag{9}$$

where $\mathbf{C}_1$ = E[$\mathbf{ss}^\mathrm{T}$], $\mathbf{C}_2$ = E[$\mathbf{tt}^\mathrm{T}$], and $\mathbf{C}_3$ = E[$\mathbf{st}^\mathrm{T}$]. These matrices can be directly obtained from $\mathbf{C}(0)$. To find the optimal readout matrix $\mathbf{B}_{\mathrm{opt}}$, we minimize ε with respect to $\mathbf{B}$, leading to the Ordinary Least Squares solution:

$$\mathbf{B}_{\mathrm{opt}} = \mathbf{C}_1^{-1}\mathbf{C}_3, \tag{10}$$

First, we define the covariance of the target activity *predicted* by the optimal model (note the "hats"), which we denote as $\hat{\mathbf{C}}_2$:

$$\hat{\mathbf{C}}_2 = \mathbb{E}\left[\hat{\mathbf{t}}\hat{\mathbf{t}}^\top\right] = \mathbf{B}_{\mathrm{opt}}^\top \mathbf{C}_1 \mathbf{B}_{\mathrm{opt}} \tag{11}$$

By substituting the expression for $\mathbf{B}_{\mathrm{opt}}$, we arrive at a key relationship:

$$\hat{\mathbf{C}}_2 = \left(\mathbf{C}_1^{-1}\mathbf{C}_3\right)^\top \mathbf{C}_1 \left(\mathbf{C}_1^{-1}\mathbf{C}_3\right) = \mathbf{C}_3^\top \mathbf{C}_1^{-1}\mathbf{C}_3 \tag{12}$$

The principal dimensions of this predicted covariance matrix, $\hat{\mathbf{C}}_2$, define the communication subspace. These are found by computing the eigen-decomposition:

$$\hat{\mathbf{C}}_2 = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^\top \tag{13}$$

Here, the columns of the orthogonal matrix $\mathbf{V}$ are the eigenvectors of $\hat{\mathbf{C}}_2$ and $\mathbf{\Lambda}$ is a diagonal matrix of the corresponding eigenvalues, sorted in descending order. The prediction performance for a reduced-rank approximation of dimension i is defined as: A rank-i approximation, $\mathbf{B}_{\mathrm{RRR}}$, is constructed by projecting the optimal solution $\mathbf{B}_{\mathrm{opt}}$ onto the subspace spanned by the top i eigenvectors of $\hat{\mathbf{C}}_2$.

$$\mathbf{B}_{\mathrm{RRR}}(i) = \mathbf{B}_{\mathrm{opt}}\,\mathbf{V}(i)\mathbf{V}^\top(i) = \mathbf{B}_{\mathrm{opt}}\,\mathbf{P}_i \tag{14}$$

where $\mathbf{V}(i)$ is a matrix containing first i columns of $\mathbf{V}$. Here $\mathbf{P}_i$ = $\mathbf{V}(i)\mathbf{V}^\mathrm{T}(i)$ is the projection matrix. Note that $\mathbf{P}_i$ is symmetric $\left(\mathbf{P}_i^\top = \mathbf{P}_i\right)$ and idempotent $\left(\mathbf{P}_i^2 = \mathbf{P}_i\right)$. We define the prediction performance as:

$$\mathrm{Performance}\,(i) = 1 - \frac{\varepsilon_i}{\varepsilon_0}, \tag{15}$$

where ε_i is the MSE for the rank-i approximation

$$\varepsilon_i = \mathrm{Tr}\left(\mathbf{C}_2 + \mathbf{B}_{\mathrm{RRR}}^\top(i)\mathbf{C}_1\mathbf{B}_{\mathrm{RRR}}(i) - 2\mathbf{B}_{\mathrm{RRR}}^\top(i)\mathbf{C}_3\right). \tag{16}$$

ε_0 is the MSE for the case where the readout matrix is a null matrix, corresponding to the total variance of the target population (ε_0 = Tr($\mathbf{C}_2$)).

To simplify the expressions for ε_i. We substitute $\mathbf{B}_{\mathrm{RRR}}(i) = \mathbf{B}_{\mathrm{opt}}\mathbf{P}_i$ into the expression:

$$\varepsilon_i = \mathrm{Tr}\left(\mathbf{C}_2\right) - \mathrm{Tr}\left(2\mathbf{P}_i\mathbf{B}_{\mathrm{opt}}^{\top}\mathbf{C}_3 - \mathbf{P}_i\mathbf{B}_{\mathrm{opt}}^{\top}\mathbf{C}_1\mathbf{B}_{\mathrm{opt}}\mathbf{P}_i\right) \tag{17}$$

Now, we substitute $\mathbf{B}_{\mathrm{opt}}^{\top} = \mathbf{C}_3^{\top}\mathbf{C}_1^{-1}$ and recognize that the term $\hat{\mathbf{C}}_2$ is $\hat{\mathbf{C}}_2$. Using the cyclic property of the trace (Tr(**AB**) = Tr(**BA**)) and the idempotent property of the projection matrix $(\mathbf{P}_i^2 = \mathbf{P}_i)$, we can simplify further:

$$\varepsilon_i = \mathrm{Tr}\left(\mathbf{C}_2\right) - \mathrm{Tr}\left(\hat{\mathbf{C}}_2\mathbf{P}_i\right) \tag{18}$$

By applying the cyclic property and expanding $\mathbf{P}_i$, we can see that the second term is the sum of the top i eigenvalues (λ_j) of $\hat{\mathbf{C}}_2$.

$$\mathrm{Tr}\left(\hat{\mathbf{C}}_2\mathbf{V}(i)\mathbf{V}(i)^{\top}\right) = \mathrm{Tr}\left(\mathbf{V}(i)^{\top}\hat{\mathbf{C}}_2\mathbf{V}(i)\right) = \sum_{j=1}^{i}\lambda_j \tag{19}$$

Now the expression for prediction performance simplifies to:

$$\mathrm{Performance}(i) = \frac{\sum_{j=1}^{i}\lambda_j}{\mathrm{Tr}\left(\mathbf{C}_2\right)}, \tag{20}$$

Which can also be interpreted as:

$$\mathrm{Performance}\ (i) = \frac{\mathrm{Explained\ variance}\ (i\ \mathrm{dimensions}\ )}{\mathrm{Total\ target\ variance}}. \tag{21}$$

The maximum prediction performance (full-rank approximation) is given by:

$$\mathrm{Performance}\ = \frac{\mathrm{Tr}\left(\hat{\mathbf{C}}_2\right)}{\mathrm{Tr}\left(\mathbf{C}_2\right)} = \frac{\mathrm{Tr}\left(\mathbf{C}_3^{\mathrm{T}}\mathbf{C}_1^{-1}\mathbf{C}_3\right)}{\mathrm{Tr}\left(\mathbf{C}_2\right)} \tag{22}$$

For the communication-subspace analyses (Figs. 2g and 3k,l), source and target populations of 30 principal neurons were drawn at random from the 72 neurons of each area, and prediction performance and dimensionality were averaged over random draws. The dependence of the dimensionality on the number of modeled and sampled neurons is characterized in Suppl. Fig. S9.

Interpretation of the results

The prediction performance (Eq. 22) is governed by the interplay between the source ($\mathbf{C}_1$), target ($\mathbf{C}_2$), and cross-population ($\mathbf{C}_3$) covariances. The effectiveness of communication (indicated by a large predictive performance) can be understood through three key factors: 1) *Source-target correlation*: Prediction performance is directly enhanced by a stronger correlation between the source and target populations (i.e., larger $\mathbf{C}_3$ norm). 2) *Signal reliability*: Prediction performance is boosted by the reliability of the signals. Less noisy activity in source and target populations, indicated by smaller diagonal entries in $\mathbf{C}_1$ and $\mathbf{C}_2$, leads to better prediction performance. 3) *Alignment*: Prediction performance is maximized when the principal directions of source covariance (eigenvectors of $\mathbf{C}_1$) align with the principal directions of source-target covariance (left singular vectors of $\mathbf{C}_3$).

This framework also explains the results in the frequency domain. The key metric here is the cross-power spectrum, which is the Fourier transform of the cross-covariances, and coherence, which is the squared cross-power spectrum normalized by the product of the auto-spectra, Eq.

7. At frequencies of high coherence, i.e., of large cross-covariance between source and target relative to their auto-covariances, the prediction performance is necessarily highest as well, since both quantities are computed from the same cross-spectral density matrix.

The dimensionality of the communication subspace is given by the rank of the matrix $\hat{\mathbf{C}}_2 = \mathbf{C}_3^{\mathrm{T}}\mathbf{C}_1^{-1}\mathbf{C}_3$. Since the source covariance matrix $\mathbf{C}_1$ is a full-rank matrix, the dimensionality of the subspace is governed by the rank of the cross-covariance matrix, $\mathbf{C}_3$. In the frequency domain, this means the dimensionality at a specific frequency is determined by the rank of the cross-power spectral density matrix (see Suppl. Fig. S5a). Our results show that the extent of normalization from surrounding neurons plays a critical role in shaping the dimensionality of the communication subspace. With surround normalization, dimensionality decreases specifically at frequencies where coherence is strongest (Fig. 4c). Normalization reduces dimensionality because when just a small subset of neurons is strongly driven, and their responses are highly correlated across areas (V1 and V2), normalization suppresses the activity of surrounding neurons with weaker drive. This suppression, resembling a soft winner-take-all process, diminishes the contribution of those less active neurons to the population's shared variability, leading to a lower effective dimensionality.

In contrast, without this normalization, no such correlation between coherence and dimensionality is observed (Fig. S4h).

Frequency decomposition of the communication subspace

To analyze how communication between different brain areas varies across frequencies, we consider the frequency decomposition of the neural responses. The Wiener-Khinchin theorem states that the autocorrelation function $\mathbf{C}(\tau)$ is the inverse Fourier transform of the power spectral density (PSD) matrix ($\boldsymbol{\mathcal{S}}(\omega)$):

$$\mathrm{C}(\tau) = \int_{-\infty}^{\infty} \mathcal{S}(\omega) e^{\imath\omega\tau} d\omega \tag{23}$$

To isolate $\mathbf{C}(\tau)$ at a particular frequency (ω^*), we apply an idealized narrow band-pass filter centered at ω^*. Specifically, the filter's transfer function $H(\omega)$ is given by,

$$H(\omega) = \begin{cases} 1 & \text{if } \omega \in A \\ 0 & \text{otherwise} \end{cases} \tag{24}$$

where $A = [\omega^* - \epsilon, \omega^* + \epsilon] \cup [-\omega^* - \epsilon, -\omega^* + \epsilon]$ and ϵ is a small positive value defining the narrow bandwidth 2ϵ around ω^*. The spectral density matrix of this filtered signal is given by,

$$\boldsymbol{\mathcal{S}}_{\text{filtered}}(\omega) = |H(\omega)|^2 \boldsymbol{\mathcal{S}}(\omega) \tag{25}$$

The resulting autocorrelation of the filtered signal is expressed as:

$$\begin{aligned} \mathrm{C}_{\text{filtered}}(\tau) &= \int_{-\infty}^{\infty} \mathcal{S}_{\text{filtered}}(\omega) e^{\imath\omega\tau} d\omega \\ &= \int_{-\infty}^{\infty} |H(\omega)|^2 \mathcal{S}(\omega) e^{\imath\omega\tau} d\omega. \end{aligned} \tag{26}$$

Using the definition of $H(\omega)$ and assuming ϵ to be sufficiently small such that (ω) and $e^{\imath\omega\tau}$ are approximately constant over the integration intervals of width 2ϵ, the frequency-specific autocorrelation at ω^* is:

$$\mathbf{C}(\tau)|_{\omega^*} \approx 2\epsilon \left(\mathcal{S}\left(-\omega^*\right) e^{-\imath\omega^*\tau} + \mathcal{S}\left(\omega^*\right) e^{\imath\omega^*\tau} \right). \tag{27}$$

At zero time lag ($\tau = 0$),

$$\mathbf{C}(0)|_{\omega^*} \approx 2\epsilon \left(\mathcal{S}\left(-\omega^*\right) + \mathcal{S}\left(\omega^*\right) \right). \tag{28}$$

Since the PSD is Hermitian, i.e., $\boldsymbol{S}(\omega) = {}^*\boldsymbol{S}(-\omega)$), we have

$$\mathbf{C}(0)|_{\omega^*} = 4\epsilon \,\mathrm{Re}\left\{ \mathcal{S}\left(\omega^*\right) \right\} \tag{29}$$

This frequency-specific autocorrelation matrix ($\mathbf{C}(0)|_{\omega}*$) can then be used in place of $\mathbf{C}(0)$ in the communication subspace analysis described above to evaluate the prediction performance at any frequency. As our analysis does not depend on the absolute scaling of this matrix, the prefactor 4ϵ can be disregarded, and we effectively use Re $\{\boldsymbol{S}(\omega^*)\}$ as the frequency-specific correlation matrix.

Application to experimental measurements

The same construction applies directly to experimental recordings, and no band-pass filtering or separate regression per frequency band is required. Given simultaneously recorded source (e.g., V1) and target (e.g., V2) population activity (spike counts or LFP), one estimates the cross-spectral density matrix $\boldsymbol{S}(\omega)$ of the pooled population (e.g., by Welch's or the multitaper method). At each frequency ω^*, Re $\{\boldsymbol{S}(\omega^*)\}$ is the frequency-specific covariance matrix; partitioning it into the source, target and cross blocks $\mathbf{C}_1$, $\mathbf{C}_2$ and $\mathbf{C}_3$ gives $\hat{\mathbf{C}}_2 = \mathbf{C}_3^{\mathrm{T}}\mathbf{C}_1^{-1}\mathbf{C}_3$, the prediction performance at each rank (Eq. 20), and the dimensionality, i.e., the number of eigenvalues of $\hat{\mathbf{C}}_2$ required to reach the full-rank performance (Eq. 22) within a fixed tolerance. Repeating this at every frequency yields the experimental counterparts of Fig. 4b,c. The prediction to be tested is that the dimensionality is smallest at the frequency of peak V1-V2 coherence and that this frequency increases with contrast.

Code availability

The code to generate all the figures can be found in the following GitHub repository: *https://github.com/asit-pal/Hierarchical-ORGaNICs*.

Data availability

The current manuscript is a computational study, so no data have been generated for this manuscript. Modelling code is uploaded as Source code.

Acknowledgements

We thank Pascal Fries and Adam Kohn for generously providing the experimental data in Fig 2. We acknowledge Michael Halassa, Hillel Adesnik, Adam Kohn and Flaviano Morone for their valuable feedback on the manuscript, and Ruben Coen-Cagli, John Maunsell, and Jason MacLean for insightful discussions. This work was initiated in part at the Aspen Center for Physics, which is supported by National Science Foundation grant PHY-2210452. This work was supported by the National Institute of Health under award numbers R01MH137669 and R01EY035242. S.M. also acknowledges support from the Simons Center for Computational Physical Chemistry (Simons Foundation grant 839534, MT). This work was supported in part through the NYU IT High-Performance Computing resources, services, and staff expertise.

Additional files

Supplementary Information

Additional information

Funding

Funder	Grant reference number	Author
National Science Foundation (NSF)	PHY-2210452	Stefano Martiniani
National Institutes of Health	R01MH137669	Stefano Martiniani
National Institutes of Health	R01EY035242	Stefano Martiniani
Simons Foundation (SF)	839534	Stefano Martiniani

Author ORCID iDs

Asit Pal: https://orcid.org/0000-0002-6654-1796
David J Heeger: https://orcid.org/0000-0002-3282-9898

References

[1] **Heeger David J** (1992) Normalization of cell responses in cat striate cortex. *Visual neuroscience* **9**:181-197 https://doi.org/10.1017/s0952523800009640 | PubMed

[2] **Baluch Farhan,** Itti Laurent (2011) Mechanisms of top-down attention. *Trends in neurosciences* **34**:210-224 https://doi.org/10.1016/j.tins.2011.02.003 | PubMed

[3] **Quax Silvan,** Jensen Ole, Tiesinga Paul (2017) Top-down control of cortical gamma-band communication via pulvinar induced phase shifts in the alpha rhythm. *PLoS computational biology* **13**:e1005519 https://doi.org/10.1371/journal.pcbi.1005519 | PubMed

[4] **Makino Hiroshi,** Komiyama Takaki (2015) Learning enhances the relative impact of top-down processing in the visual cortex. *Nature neuroscience* **18**:1116-1122 https://doi.org/10.1038/nn.4061 | PubMed

[5] **Heeger David J** (2017) Theory of cortical function. *Proceedings of the National Academy of Sciences* **114**:1773-1782 https://doi.org/10.1073/pnas.1619788114 | PubMed

[6] **Heeger David J,** Zemlianova Klavdia O (2020) A recurrent circuit implements normalization, simulating the dynamics of V1 activity. *Proceedings of the National Academy of Sciences* **117**:22494-22505 https://doi.org/10.1073/pnas.2005417117 | PubMed

[7] **Heeger David J,** Mackey Wayne E (2019) Oscillatory recurrent gated neural integrator circuits (ORGaNICs), a unifying theoretical framework for neural dynamics. *Proceedings of the National Academy of Sciences* **116**:22783-22794 https://doi.org/10.1073/pnas.1911633116 | PubMed

[8] **Rawat Shivang,** Heeger David, Martiniani Stefano (2024) Unconditional stability of a recurrent neural circuit implementing divisive normalization. In: Advances in Neural Information Processing Systems. **37** pp. 14712-14750 https://doi.org/10.52202/079017-0470 | PubMed

[9] **Morone Flaviano,** et al. (2025) Stabilization of recurrent neural networks through divisive normalization. *bioRxiv* https://doi.org/10.1101/2025.05.16.654567 | PubMed

[10] **Fries Pascal** (2005) A mechanism for cognitive dynamics: neuronal communication through neuronal coherence. *Trends in cognitive sciences* **9**:474-480 https://doi.org/10.1016/j.tics.2005.08.011 | PubMed

[11] **Salinas Emilio,** Sejnowski Terrence J (2001) Correlated neuronal activity and the flow of neural information. *Nature reviews neuroscience* **2**:539-550 https://doi.org/10.1038/35086012 | PubMed

[12] **Varela Francisco,** et al. (2001) The brainweb: phase synchronization and large-scale integration. *Nature reviews neuroscience* **2**:229-239 https://doi.org/10.1038/35067550 | PubMed

[13] **Schneider Marius,** et al. (2021) A mechanism for interareal coherence through communication based on connectivity and oscillatory power. *Neuron* **109**:4050-4067 https://doi.org/10.1016/j.neuron.2021.09.037 | PubMed

[14] **Ray Supratim,** Maunsell John HR (2015) Do gamma oscillations play a role in cerebral cortex?. *Trends in cognitive sciences* **19**:78-85 https://doi.org/10.1016/j.tics.2014.12.002 | PubMed

[15] **Semedo João D,** et al. (2019) Cortical areas interact through a communication subspace. *Neuron* **102**:249-259 https://doi.org/10.1016/j.neuron.2019.01.026 | PubMed

[16] **Poggi Peter G,** Mihalas Stefan, Mastrovito Dana (2024) Communication subspaces align with training in ANNs. *bioRxiv* https://doi.org/10.1101/2024.11.11.623065

[17] **Barbosa Joao,** et al. (2023) Early selection of task-relevant features through population gating. *Nature Communications* **14**:6837 https://doi.org/10.1038/s41467-023-42519-5 | PubMed

[18] **Gozel Olivia,** Doiron Brent (2024) Between-area communication through the lens of within-area neuronal dynamics. *Science Advances* **10**:eadl6120 https://doi.org/10.1126/sciadv.adl6120 | PubMed

[19] **Clark David G,** Beiran Manuel (2025) Structure of activity in multiregion recurrent neural networks. *Proceedings of the National Academy of Sciences* **122**:e2404039122 https://doi.org/10.1073/pnas.2404039122 | PubMed

[20] **Mastrogiuseppe Francesca,** et al. (2026) Input-dependent directionality of interactions between cortical areas. *bioRxiv* https://doi.org/10.64898/2026.05.18.725829 | PubMed

[21] **Carandini Matteo,** Heeger David J (2012) Normalization as a canonical neural computation. *Nature reviews neuroscience* **13**:51-62 https://doi.org/10.1038/nrn3136 | PubMed

[22] **Schwartz Odelia,** Simoncelli Eero P (2001) Natural signal statistics and sensory gain control. *Nature neuroscience* **4**:819-825 https://doi.org/10.1038/90526 | PubMed

[23] **Olsen Shawn R,** Bhandawat Vikas, Wilson Rachel I (2010) Divisive normalization in olfactory population codes. *Neuron* **66**:287-299 https://doi.org/10.1016/j.neuron.2010.04.009 | PubMed

[24] **McClain Kathryn,** et al. (2026) A minimal mechanism to generate long timescales without fine tuning. *arXiv* https://doi.org/10.48550/arxiv.2608.21586

[25] **Adesnik Hillel** (2017) Synaptic Mechanisms of Feature Coding in the Visual Cortex of Awake Mice. *Neuron* **95**:1147-1159 https://doi.org/10.1016/j.neuron.2017.08.014 | PubMed

[26] **Ozeki Hirofumi,** et al. (2009) Inhibitory stabilization of the cortical network underlies visual surround suppression. *Neuron* **62**:578-592 https://doi.org/10.1016/j.neuron.2009.03.028 | PubMed

[27] **Douglas Rodney J,** et al. (1995) Recurrent excitation in neocortical circuits. *Science* **269**:981-985 https://doi.org/10.1126/science.7638624 | PubMed

[28] **Douglas Rodney J,** Martin Kevan AC (2007) Recurrent neuronal circuits in the neocortex. *Current biology* **17**:R496-R500 https://doi.org/10.1016/j.cub.2007.04.024 | PubMed

[29] **Levitt Jonathan B,** Kiper Daniel C, Movshon J Anthony (1994) Receptive fields and functional architecture of macaque V2. *Journal of neurophysiology* **71**:2517-2542 https://doi.org/10.1152/jn.1994.71.6.2517 | PubMed

[30] **Geisler Wilson S,** Albrecht Duane G (1997) Visual cortex neurons in monkeys and cats: detection, discrimination, and identification. *Visual neuroscience* **14**:897-919 https://doi.org/10.1017/s0952523800011627 | PubMed

[31] **Roberts Mark J,** et al. (2013) Robust gamma coherence between macaque V1 and V2 by dynamic frequency matching. *Neuron* **78**:523-536 https://doi.org/10.1016/j.neuron.2013.03.003 | PubMed

[32] **Sclar Gary,** Maunsell John HR, Lennie Peter (1990) Coding of image contrast in central visual pathways of the macaque monkey. *Vision research* **30**:1-10 https://doi.org/10.1016/0042-6989(90)90123-3 | PubMed

[33] **Buzsaki Gyorgy,** Draguhn Andreas (2004) Neuronal oscillations in cortical networks. *Science* **304**:1926-1929 https://doi.org/10.1126/science.1099745 | PubMed

[34] **Buzsáki György,** Anastassiou Costas A, Koch Christof (2012) The origin of extracellular fields and currents—EEG, ECoG, LFP and spikes. *Nature reviews neuroscience* **13**:407-420 https://doi.org/10.1038/nrn3241 | PubMed

[35] **Ray Supratim,** Ni Amy M, Maunsell John HR (2013) Strength of gamma rhythm depends on normalization. *PLoS Biology* **11**:e1001477 https://doi.org/10.1371/journal.pbio.1001477 | PubMed

[36] **Hermes Dora,** et al. (2019) An image-computable model for the stimulus selectivity of gamma oscillations. *eLife* **8**:e47035 https://doi.org/10.7554/eLife.47035 | PubMed

[37] **Miller Kai J,** et al. (2009) Power-law scaling in the brain surface electric potential. *PLoS computational biology* **5**:e1000609 https://doi.org/10.1371/journal.pcbi.1000609 | PubMed

[38] **Engel Andreas K,** Fries Pascal, Singer Wolf (2001) Dynamic predictions: oscillations and synchrony in top–down processing. *Nature Reviews Neuroscience* **2**:704-716 https://doi.org/10.1038/35094565 | PubMed

[39] **Fries Pascal** (2015) Rhythms for cognition: communication through coherence. *Neuron* **88**:220-235 https://doi.org/10.1016/j.neuron.2015.09.034 | PubMed

[40] **Srinath Ramanujan,** Ruff Douglas A, Cohen Marlene R (2021) Attention improves information flow between neuronal populations without changing the communication subspace. *Current Biology* **31**:5299-5313 https://doi.org/10.1016/j.cub.2021.09.076 | PubMed

[41] **Hutchison R Matthew,** et al. (2013) Dynamic functional connectivity: promise, issues, and interpretations. *Neuroimage* **80**:360-378 https://doi.org/10.1016/j.neuroimage.2013.05.079 | PubMed

[42] **Allen Elena A,** et al. (2014) Tracking whole-brain connectivity dynamics in the resting state. *Cerebral cortex* **24**:663-676 https://doi.org/10.1093/cercor/bhs352 | PubMed

[43] **Mishkin Mortimer,** Ungerleider Leslie G, Macko Kathleen A (1983) Object vision and spatial vision: two cortical pathways. *Trends in neurosciences* **6**:414-417 https://doi.org/10.1016/0166-2236(83)90190-x

[44] **Maunsell John HR,** Treue Stefan (2006) Feature-based attention in visual cortex. *Trends in neurosciences* **29**:317-322 https://doi.org/10.1016/j.tins.2006.04.001 | PubMed

[45] **Saproo Sameer,** Serences John T (2014) Attention improves transfer of motion information between V1 and MT. *Journal of Neuroscience* **34**:3586-3596 https://doi.org/10.1523/jneurosci.3484-13.2014 | PubMed

[46] **Hochreiter Sepp,** Schmidhuber Jürgen (1997) Long short-term memory. *Neural computation* **9**:1735-1780 https://doi.org/10.1162/neco.1997.9.8.1735 | PubMed

[47] **Sutskever Ilya,** Vinyals Oriol, Le Quoc V (2014) Sequence to sequence learning with neural networks. In: Advances in neural information processing systems. **27**

[48] **Rao Rajesh PN,** Ballard Dana H (1997) Dynamic model of visual recognition predicts neural response properties in the visual cortex. *Neural computation* **9**:721-763 https://doi.org/10.1162/neco.1997.9.4.721 | PubMed

[49] **Wacongne Catherine,** Changeux Jean-Pierre, Dehaene Stanislas (2012) A neuronal model of predictive coding accounting for the mismatch negativity. *Journal of Neuroscience* **32**:3665-3678 https://doi.org/10.1523/jneurosci.5003-11.2012 | PubMed

[50] **Rao Rajesh PN,** Ballard Dana H (1999) Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature neuroscience* **2**:79-87 https://doi.org/10.1038/4580 | PubMed

[51] **Lee Tai Sing,** Mumford David (2003) Hierarchical Bayesian inference in the visual cortex. *Journal of the Optical Society of America A* **20**:1434-1448 https://doi.org/10.1364/josaa.20.001434 | PubMed

[52] **Friston Karl** (2005) A theory of cortical responses. *Philosophical transactions of the Royal Society B: Biological sciences* **360**:815-836 https://doi.org/10.1098/rstb.2005.1622 | PubMed

[53] **Bastos Andre M,** et al. (2012) Canonical microcircuits for predictive coding. *Neuron* **76**:695-711 https://doi.org/10.1016/j.neuron.2012.10.038 | PubMed

[54] **Burns Samuel P,** Xing Dajun, Shapley Robert M (2010) Comparisons of the dynamics of local field potential and multiunit activity signals in macaque visual cortex. *Journal of Neuroscience* **30**:13739-13749 https://doi.org/10.1523/jneurosci.0743-10.2010 | PubMed

[55] **Carandini Matteo,** Heeger David J, Movshon J Anthony (1997) Linearity and normalization in simple cells of the macaque primary visual cortex. *Journal of Neuroscience* **17**:8621-8644 https://doi.org/10.1523/jneurosci.17-21-08621.1997 | PubMed

[56] **Rubin Daniel B,** Van Hooser Stephen D, Miller Kenneth D (2015) The stabilized supralinear network: a unifying circuit motif underlying multi-input integration in sensory cortex. *Neuron* **85**:402-417 https://doi.org/10.1016/j.neuron.2014.12.026 | PubMed

[57] **Angelucci Alessandra,** Levitt Jonathan B, Lund Jennifer S (2002) Anatomical origins of the classical receptive field and modulatory surround field of single neurons in macaque visual cortical area V1. *Progress in brain research* **136**:373-388 https://doi.org/10.1016/s0079-6123(02)36031-x | PubMed

[58] **Ma Wei Ji,** et al. (2006) Bayesian inference with probabilistic population codes. *Nature neuroscience* **9**:1432-1438 https://doi.org/10.1038/nn1790 | PubMed

[59] **Knill David C,** Pouget Alexandre (2004) The Bayesian brain: the role of uncertainty in neural coding and computation. *TRENDS in Neurosciences* **27**:712-719 https://doi.org/10.1016/j.tins.2004.10.007 | PubMed

[60] **Ni Shencong,** Harris Brendan, Gong Pulin (2024) Distributed and dynamical communication: a mechanism for flexible cortico-cortical interactions and its functional roles in visual attention. *Communications Biology* **7**:550 https://doi.org/10.1038/s42003-024-06228-z | PubMed

[61] **Honey Christopher J,** et al. (2009) Predicting human resting-state functional connectivity from structural connectivity. *Proceedings of the National Academy of Sciences* **106**:2035-2040 https://doi.org/10.1073/pnas.0811168106 | PubMed

[62] **Treue Stefan,** Trujillo Julio C Martinez (1999) Feature-based attention influences motion processing gain in macaque visual cortex. *Nature* **399**:575-579 https://doi.org/10.1038/21176 | PubMed

[63] **Saenz Melissa,** Buracas Giedrius T, Boynton Geoffrey M (2002) Global effects of feature-based attention in human visual cortex. *Nature neuroscience* **5**:631-632 https://doi.org/10.1038/nn876 | PubMed

[64] **Reynolds John H,** Heeger David J (2009) The normalization model of attention. *Neuron* **61**:168-185 https://doi.org/10.1016/j.neuron.2009.01.002 | PubMed

[65] **Hughes Stuart W,** Crunelli Vincenzo (2005) Thalamic mechanisms of EEG alpha rhythms and their pathological implications. *The Neuroscientist* **11**:357-372 https://doi.org/10.1177/1073858405277450 | PubMed

[66] **Bollimunta Anil,** et al. (2011) Neuronal mechanisms and attentional modulation of corticothalamic alpha oscillations. *Journal of Neuroscience* **31**:4935-4943 https://doi.org/10.1523/jneurosci.5580-10.2011 | PubMed

[67] **Halgren Mila,** et al. (2019) The generation and propagation of the human alpha rhythm. *Proceedings of the National Academy of Sciences* **116**:23772-23782 https://doi.org/10.1073/pnas.1913092116 | PubMed

[68] **Jensen Ole,** Mazaheri Ali (2010) Shaping functional architecture by oscillatory alpha activity: gating by inhibition. *Frontiers in human neuroscience* **4**:186 https://doi.org/10.3389/fnhum.2010.00186 | PubMed

[69] **Klimesch Wolfgang,** Sauseng Paul, Hanslmayr Simon (2007) EEG alpha oscillations: the inhibition–timing hypothesis. *Brain research reviews* **53**:63-88 https://doi.org/10.1016/j.brainresrev.2006.06.003 | PubMed

[70] **Schmitt L Ian,** et al. (2017) Thalamic amplification of cortical connectivity sustains attentional control. *Nature* **545**:219-223 https://doi.org/10.1038/nature22073 | PubMed

[71] **Sherman S Murray,** Usrey W Martin (2021) *Exploring Thalamocortical Interactions: Circuitry for Sensation, Action, and Cognition* https://doi.org/10.1093/med/9780197503874.001.0001

[72] **Schmolesky Matthew T,** et al. (1998) Signal timing across the macaque visual system. *Journal of neurophysiology* **79**:3272-3278 https://doi.org/10.1152/jn.1998.79.6.3272 | PubMed

[73] **Semedo João D,** et al. (2022) Feedforward and feedback interactions between visual cortical areas use different population activity patterns. *Nature communications* **13**:1099 https://doi.org/10.1038/s41467-022-28552-w | PubMed

[74] **Bekkers John M** (2011) Pyramidal neurons. *Current biology* **21**:R975 https://doi.org/10.1016/j.cub.2011.10.037 | PubMed

[75] **Riedemann Therese** (2019) Diversity and function of somatostatin-expressing interneurons in the cerebral cortex. *International journal of molecular sciences* **20**:2952 https://doi.org/10.3390/ijms20122952 | PubMed

[76] **Adesnik Hillel,** et al. (2012) A neural circuit for spatial summation in visual cortex. *Nature* **490**:226-231 https://doi.org/10.1038/nature11526 | PubMed

[77] **Pfeffer Carsten K,** et al. (2013) Inhibition of inhibition in visual cortex: the logic of connections between molecularly distinct interneurons. *Nature neuroscience* **16**:1068-1076 https://doi.org/10.1038/nn.3446 | PubMed

[78] **Campagnola Luke,** et al. (2022) Local connectivity and synaptic dynamics in mouse and human neocortex. *Science* **375**:eabj5861 https://doi.org/10.1126/science.abj5861 | PubMed

[79] **Lee Jung Hoon,** Mihalas Stefan (2017) Visual processing mode switching regulated by VIP cells. *Scientific Reports* **7**:1843 https://doi.org/10.1038/s41598-017-01830-0 | PubMed

[80] **Atallah Bassam V,** et al. (2012) Parvalbumin-expressing interneurons linearly transform cortical responses to visual stimuli. *Neuron* **73**:159-170 https://doi.org/10.1016/j.neuron.2011.12.013 | PubMed

[81] **Sanzeni Alessandro,** et al. (2020) Inhibition stabilization is a widespread property of cortical networks. *eLife* **9**:e54875 https://doi.org/10.7554/eLife.54875 | PubMed

[82] **Studer Florian,** Zeppillo Tommaso, Barkat Tania Rinaldi (2024) Top-down inputs are controlled by somatostatin-expressing interneurons during associative learning. *bioRxiv* https://doi.org/10.1101/2024.12.18.629098

[83] **Naka Alexander,** et al. (2019) Complementary networks of cortical somatostatin interneurons enforce layer specific control. *eLife* **8**:e43696 https://doi.org/10.7554/eLife.43696 | PubMed

[84] **Shen Shan,** et al. (2022) Distinct organization of two cortico-cortical feedback pathways. *Nature communications* **13**:6389 https://doi.org/10.1038/s41467-022-33883-9 | PubMed

[85] **Zhu J Julius** (2000) Maturation of layer 5 neocortical pyramidal neurons: amplifying salient layer 1 and layer 4 inputs by Ca2+ action potentials in adult rat tuft dendrites. *The Journal of physiology* **526**:571-587 https://doi.org/10.1111/j.1469-7793.2000.00571.x | PubMed

[86] **Carandini Matteo,** Heeger David J (1994) Summation and division by neurons in primate visual cortex. *Science* **264**:1333-1336 https://doi.org/10.1126/science.8191289 | PubMed

[87] **Chance Frances S,** Abbott Larry F, Reyes Alex D (2002) Gain modulation from background synaptic input. *Neuron* **35**:773-782 https://doi.org/10.1016/s0896-6273(02)00820-6 | PubMed

[88] **Ferguson Katie A,** Cardin Jessica A (2020) Mechanisms underlying gain modulation in the cortex. *Nature Reviews Neuroscience* **21**:80-92 https://doi.org/10.1038/s41583-019-0253-y | PubMed

[89] **Srinivasan Mandyam V,** Bernard Gary D (1976) A proposed mechanism for multiplication of neural signals. *Biological Cybernetics* **21**:227-236 https://doi.org/10.1007/bf00344168 | PubMed

[90] **Polsky Alon,** Mel Bartlett W, Schiller Jackie (2004) Computational subunits in thin dendrites of pyramidal cells. *Nature Neuroscience* **7**:621-627 https://doi.org/10.1038/nn1253 | PubMed

[91] **Silver R Angus** (2010) Neuronal arithmetic. *Nature Reviews Neuroscience* **11**:474-489 https://doi.org/10.1038/nrn2864 | PubMed

[92] **Corbetta Maurizio,** et al. (1991) Selective and divided attention during visual discriminations of shape, color, and speed: functional anatomy by positron emission tomography. *Journal of neuroscience* **11**:2383-2402 https://doi.org/10.1523/jneurosci.11-08-02383.1991 | PubMed

[93] **Huk Alexander C,** Heeger David J (2000) Taskrelated modulation of visual cortex. *Journal of neurophysiology* **83**:3525-3536 https://doi.org/10.1152/jn.2000.83.6.3525 | PubMed

[94] **Rawat Shivang,** Heeger David, Martiniani Stefano (2024) A Comprehensive Large-Scale Model of the Primary Visual Cortex (V1). *Bulletin of the American Physical Society*

[95] **Carandini Matteo** (2004) Amplification of trial-totrial response variability by neurons in visual cortex. *PLoS biology* **2**:e264 https://doi.org/10.1371/journal.pbio.0020264 | PubMed

[96] **Rawat Shivang,** Martiniani Stefano (2024) Element-wise and recursive solutions for the power spectral density of biological stochastic dynamical systems at fixed points. *Physical Review Research* **6**:043179 https://doi.org/10.1103/physrevresearch.6.043179 | PubMed

Peer reviews

Reviewer #1 (Public review):

[Editors' note: this version has been assessed by the Reviewing Editor without further input from the original reviewers. The authors have addressed the comments raised in the previous round of review.]

In this paper, Pal and colleagues propose a mechanistic unification of two influential accounts of inter-areal communication: communication through coherence and communication subspaces. A major strength of the paper is that it does not treat coherence and communication subspaces as independent phenomena, as typically done, but instead derives both from the same circuit with divisive normalization. In this framework, noise-driven fluctuations around the normalized fixed point determine covariance and cross-power structure (which, in retrospect, makes so much sense to be related). Then, they show how these determine linear prediction performance and the effective dimensionality of the communication subspace. They also show (however not very visually, see recommendation below for a figure) how divisive normalization is crucial to shape inter-areal coherence and the dimensionality of communication.

I found this conceptual contribution potentially very influential, but somewhat obscured by the technical complexity of the model. The central intuition (I think) is that recurrent normalization can organize cross-area fluctuations, both frequency-specific correlations and cross-covariances. Took me a while to grasp this insight, mostly because I was stuck with the model details. Note that I have some experience with network dynamics, but not with this particular model.

https://doi.org/10.7554/eLife.111297.2.sa3

Reviewer #2 (Public review):

Summary:

The authors extend the ORGaNICs framework (a recurrent circuit that dynamically implements divisive normalization) to connected cortical areas with explicit top-down feedback. Because the network has a known analytical fixed point that coincides with (or closely approximates) the normalization equation, the authors can linearize about that fixed point and derive closed-form expressions for the power spectral density, inter-areal coherence, and communication subspaces. Using a two-area instantiation (V1 & V2) with a single fixed parameter set and no data fitting, they show the model reproduces: (i) contrast-response functions with steeper slope V2; (ii) gamma-band power and coherence peaks that shift to higher frequency with contrast; and (iii) a low-dimensional inter-areal communication subspace that is lower-dimensional than the within-area subspace. They derive parallel predictions of what happens by changing model parameters: feedback gain

enhances inter-areal and suppresses within-area communication, and normalization is necessary for both the oscillatory dynamics and the reduced subspace dimensionality. A three-area extension (V1&V4, V1&V5/MT) is used to argue that differential top-down feedback can dynamically route functional connectivity.

Strengths:

(1) Analytical tractability: Deriving power spectra, coherence, and communication-subspace structure in closed form from a known fixed point is genuinely valuable.

(2) Conceptual unification: Framing coherence and communication subspaces as arising from the same normalization-driven dynamics is an elegant and useful contribution.

(3) Breadth from few assumptions: A large range of phenomena (contrast gain, gamma dynamics) emerges from normalization-based model assumptions.

(4) Biological grounding: The mapping of model variables onto identified cell types connects the abstract computation to known cortical microcircuitry.

(5) The prediction that input-gain versus feedback-gain modulation produce distinct spectral signatures gives experimentalists a clear way to test the framework.

https://doi.org/10.7554/eLife.111297.2.sa2

Reviewer #3 (Public review):

Summary

The work of Pal and colleagues considers a hierarchical and multi-population version of the "oscillatory recurrent gated neural integrator circuits" (ORGaNICs) model, showing through analytics that the model captures multiple relevant experimental results: first of all, its oscillatory dynamics produce a profile with high resemblance to experimental results, both in terms of decay of power at high frequency and in terms of shifting peak as a function of stimulus contrast. Second, inter-areal communication subspace dimensionality is lower than within-area dimensionality. The authors then proceed to further characterize the model's response properties as a function of input and feedback gain. In particular, they find that frequencies transmitted with higher strength also carry more information, that changing gain modifies the dimensionality of communication subspaces, and that these properties can be used in a three-layer model, where an upstream area can select which downstream area to communicate to, based on the strength of feedback gain.

Strengths

This work demonstrates that a single-circuit model with normalization properties can capture both the oscillatory dynamics and the inter-areal communication properties measured in cortical circuits, matching multiple experimental results. The full analytical tractability of the model is highly advantageous, allowing for easier exploration of parameters, replicability, and effective interpretations of results compared to purely numerical approaches.

The work also makes a useful conceptual link between normalization, coherence-based communication, and subspace-based communication. In particular, it shows how both phenomena can emerge from the same circuit dynamics, where normalization is a key factor.

Interestingly, the model is also extended to multiple areas, showing how attention (in the form of changes in feedback gain) can synchronize the activity of a downstream area with one of two upstream areas, thus effectively selecting which area to communicate with.

In general, this is an interesting computational framework and a useful starting point for future modeling work. A particular strength is that it connects normalization, oscillatory dynamics, coherence, and communication subspaces within one analytically tractable model, making it possible to generate mechanistic hypotheses about when inter-areal communication should be stronger, lower-dimensional, or preferentially routed through feedback.

https://doi.org/10.7554/eLife.111297.2.sa1

Author response:

The following is the authors' response to the original reviews.

> ***eLife Assessment:***
>
> *This important study extends a model of cortical normalization (ORGaNICs) to interacting cortical areas and shows that communication through coherence and communication subspaces can arise from a single set of dynamics. The evidence is solid, showing analytically that contrast-dependent gamma dynamics and a low-dimensional inter-areal communication subspace arise from one parameter set, though the comparisons to data remain qualitative and the analytics rest on a linearization that is not checked against numerical simulation. The work will interest neuroscientists and theorists concerned with inter-areal communication, cortical oscillations, and divisive normalization.*

We thank the editors and reviewers for this assessment. We have addressed both stated qualifications. On the linearization, we now validate the closed-form power and coherence spectra against direct numerical simulation of the full nonlinear stochastic circuit at every contrast (new SI Fig. S6 and SI Sec. II). On the comparison to data, in addition to the unfitted comparison of Fig. 2, we now provide a quantitative comparison in which the model is fitted to the experimental power and coherence spectra, with residuals reported at every contrast (new SI Fig. S7 and SI Sec. III).

> ***Public Reviews:***
>
> ***Reviewer #1 (Public review):***
>
> *In this paper, Pal and colleagues propose a mechanistic unification of two influential accounts of inter-areal communication: communication through coherence and communication subspaces. A major strength of the paper is that it does not treat coherence and communication subspaces as independent phenomena, as typically done, but instead derives both from the same circuit with divisive normalization. In this framework, noise-driven fluctuations around the normalized fixed point determine covariance and cross-power structure (which, in retrospect, makes so much sense to be related). Then, they show how these determine linear prediction performance and the effective dimensionality of the communication subspace. They also show (however not very visually, see recommendation below for a figure) how divisive normalization is crucial to shape inter-areal coherence and the dimensionality of communication.*
>
> *I found this conceptual contribution potentially very influential, but somewhat obscured by the technical complexity of the model. The central intuition (I think) is that recurrent normalization can organize cross-area fluctuations, both frequency-specific correlations and cross-covariances. Took me a while to grasp this insight, mostly because I was stuck with the model details. Note that I have some experience with network dynamics, but not with this particular model.*

We thank the reviewer for this positive assessment. We agree that recurrent normalization shapes the cross-spectral density of the noise-driven fluctuations about the normalized fixed

point, and that power, coherence, and the communication subspace can therefore be viewed as complementary readouts of the same underlying dynamics. To make this central connection explicit from the outset, we have added a new paragraph, "The core mechanism in brief" (Results, end of "Theory"), which states this mechanism before the main results are presented. The dependence of coherence and communication-subspace dimensionality on normalization is shown in SI Figs. S4e–h and S5, which are now cited wherever the role of normalization is discussed. In addition, the new Discussion subsection "Experimental predictions of the model" describes how normalization could be manipulated experimentally and the readouts expected.

> ***Reviewer #2 (Public review):***
>
> *Summary:*
>
> *The authors extend the ORGaNICs framework (a recurrent circuit that dynamically implements divisive normalization) to connected cortical areas with explicit top-down feedback. Because the network has a known analytical fixed point that coincides with (or closely approximates) the normalization equation, the authors can linearize about that fixed point and derive closed-form expressions for the power spectral density, inter-areal coherence, and communication subspaces. Using a two-area instantiation (V1 & V2) with a single fixed parameter set and no data fitting, they show the model reproduces: (i) contrast-response functions with steeper slope V2; (ii) gamma-band power and coherence peaks that shift to higher frequency with contrast; and (iii) a low-dimensional inter-areal communication subspace that is lower-dimensional than the within-area subspace. They derive parallel predictions of what happens by changing model parameters: feedback gain enhances inter-areal and suppresses within-area communication, and normalization is necessary for both the oscillatory dynamics and the reduced subspace dimensionality. A three-area extension (V1&V4, V1&V5/MT) is used to argue that differential top-down feedback can dynamically route functional connectivity.*

We thank the reviewer for this summary.

> *Strengths:*
>
> *(1) Analytical tractability: Deriving power spectra, coherence, and communication-subspace structure in closed form from a known fixed point is genuinely valuable.*
>
> *(2) Conceptual unification: Framing coherence and communication subspaces as arising from the same normalization-driven dynamics is an elegant and useful contribution.*
>
> *(3) Breadth from few assumptions: A large range of phenomena (contrast gain, gamma dynamics) emerges from normalization-based model assumptions.*
>
> *(4) Biological grounding: The mapping of model variables onto identified cell types connects the abstract computation to known cortical microcircuitry.*
>
> *(5) The prediction that input-gain versus feedback-gain modulation produce distinct spectral signatures gives experimentalists a clear way to test the framework.*

We thank the reviewer for the positive assessment.

> *Weaknesses:*
>
> *(1) Comparisons are qualitative, not quantitative: The theory/experiment panels are visual side-by-side comparisons. There is no quantitative goodness-of-fit for any predictions.*

We thank the reviewer for this comment. We agree, and have added a quantitative comparison. New SI Fig. S7 and SI Sec. III fit the model parameters to the experimental power and coherence spectra and report the residuals at every contrast.

> *(2) The simulations use $\tau \approx 1$ms for all cell types, which the authors acknowledge is unrealistically short; realistic values would shift the gamma peaks to lower frequencies.*

We thank the reviewer for this comment. We agree. The Limitations (Discussion) state that more realistic values would shift the gamma-band oscillations to lower frequencies, and that when all time constants were released as free parameters in the fit to the experimental spectra (SI Fig. S7) they converged to values between 0.2 and 3ms. The time constants are now described consistently as effective population time constants rather than as intrinsic single-neuron time constants (Limitations).

> *(3) Divisive normalization is a special case and is recovered exactly only for the identity recurrent matrix (self-normalization). Some statements that the circuit implements divisive normalization exactly need softening.*

We thank the reviewer for this comment. We agree, and have softened the statements throughout. The main text now states that exact recovery of Eq. 2 requires $W_{11} = I$ and $\beta = \gamma = 1$, that for the centre-surround W11 used in all reported results the circuit closely approximates normalization (to within 0.6%), and that the fixed point is obtained numerically, at first mention (Results, "Theory"), in the Methods ("The Model" and the new subsection "Numerical computation of fixed points") and in the Discussion ("Normalization").

> *(4) The element-wise (multiplicative) interaction in the modulator dynamics is not tied to a specific cellular mechanism.*

We thank the reviewer for this comment. We agree: The Limitations (Discussion) state that, while weighted sums and gain control may be implemented via balanced excitatory–inhibitory mechanisms, we can only speculate about the cellular mechanisms for the element-wise products. We speculate that the element-wise product between presynaptic firing rates in Eq. 1b might be mediated by coincident spike detection (assuming that each of the spike trains is Poisson and that the synaptic current is proportional to the probability of coincident spikes within a short time window (Srinivasan & Bernard, 1976)), and that the element-wise product between postsynaptic membrane potential and presynaptic firing rate in Eq. 1c might be mediated by NMDA receptors (Polsky, Mel & Schiller, 2004; Silver, 2010) (see Discussion, last paragraph of "Biological Plausibility").

> ***Reviewer #3 (Public review):***
>
> *Summary:*
>
> *The work of Pal and colleagues considers a hierarchical and multi-population version of the "oscillatory recurrent gated neural integrator circuits" (ORGaNICs) model, showing through analytics that the model captures multiple relevant experimental results: first of all, its oscillatory dynamics produce a profile with high resemblance to experimental results, both in terms of decay of power at high frequency and in terms of shifting peak as a function of stimulus contrast. Second, inter-areal communication subspace dimensionality is lower than within-area dimensionality. The authors then proceed to further characterize the model's response properties as a function of input and feedback gain. In particular, they find that frequencies transmitted with higher strength also carry more information, that changing gain modifies the dimensionality of communication subspaces, and that these properties can be used in a three-layer model, where an upstream area can select which downstream area to communicate to, based on the strength of feedback gain.*

We thank the reviewer for the summary.

> *Strengths:*
>
> *This work demonstrates that a single-circuit model with normalization properties can capture both the oscillatory dynamics and the inter-areal communication properties measured in cortical circuits, matching multiple experimental results. The full analytical tractability of the model is highly advantageous, allowing for easier exploration of parameters, replicability, and effective interpretations of results compared to purely numerical approaches.*
>
> *The work also makes a useful conceptual link between normalization, coherence-based communication, and subspace-based communication. In particular, it shows how both phenomena can emerge from the same circuit dynamics, where normalization is a key factor.*
>
> *Interestingly, the model is also extended to multiple areas, showing how attention (in the form of changes in feedback gain) can synchronize the activity of a downstream area with one of two upstream areas, thus effectively selecting which area to communicate with.*
>
> *In general, this is an interesting computational framework and a useful starting point for future modeling work. A particular strength is that it connects normalization, oscillatory dynamics, coherence, and communication subspaces within one analytically tractable model, making it possible to generate mechanistic hypotheses about when inter-areal communication should be stronger, lower-dimensional, or preferentially routed through feedback.*

We thank the reviewer for the positive assessment.

> *Weaknesses:*
>
> *Although I see the analytic approach as a strength, at the same time I regard the lack of any numerical comparison as a big weakness. Circuit simulations would not only confirm the correctness of the analytics, but also offer further insights on the error margins and on the regimes where the analytics are valid. This is because, to my understanding, the analytics are based on a linear approximation around the operating regime, which means deviations might be expected, especially for high gain levels in the input, or in the feedforward and feedback pathways.*

We thank the reviewer for this comment. We agree that a numerical check of the linearization is required, and have added one. New SI Fig. S6 and SI Sec. II integrate the full nonlinear stochastic system (SI Eqs. S31–S44, with the firing-rate noise and the filtered synaptic noise of SI Sec. VI) with the Euler–Maruyama scheme (time step 10 μs; 100 independent trials of 16s at each contrast) and compare the V1 power spectra and V1–V2 coherence spectra estimated from the simulated trajectories (Welch's method) with the closed-form expressions at all eight contrasts of Fig. 2. At the baseline parameters, the root-mean-square deviation is below 0.017 for the normalised power and below 0.018 for the coherence, and the maximum absolute deviation is below 0.072, at every contrast; residuals are shown in the lower panels of Fig. S6. The Methods ("Power and coherence spectra") now describe this validation. On the regime of validity: the linearization holds for small fluctuations about the fixed point, as stated in the Methods, and the simulations bear this out. When the amplitudes of the noise sources are increased, the deviation between the simulated and the analytical spectra grows, and it does so first at the lowest contrasts. This is because at low contrast < 10% the activities lie near zero, and the noise fluctuations can drive them negative, where the rectification in the model clips them to zero; the linearized dynamics do

not capture this clipping. Note that changing the strength of the noise does not alter the analytical results, since the plotted spectra are all normalized. The deviations do not, however, grow with the gains: changing the input or feedback gain changes the fixed point, and the linearization about the new fixed point remains valid, because its accuracy is controlled by the strength of the fluctuations relative to the fixed point, not by the gains themselves. This characterization of the validity regime is now stated in SI Sec. II. All results reported in the paper use the baseline parameters, for which the linearization is accurate at every contrast and for all reported gain values. The stable ranges of the gains are given in SI Table S2.

> *Another problem is that the analytically tractable model seems to rely on effective connectivity weights that break Dale's law. Numerical simulations with explicitly modeled excitatory and inhibitory units might give insights into effects due, e.g., to the additional transmission delays mentioned in the Discussion.*

We thank the reviewer for this comment. We agree, and have made the distinction between effective and anatomical connectivity explicit. The Results ("Theory") and the Methods ("The Model") now state, where the matrices are introduced, that the recurrent, feedforward and feedback matrices are effective (net) connectivity, with negative entries standing for disynaptic inhibition through interneurons that are not modelled explicitly. The Limitations (Discussion) now state that the circuit as written does not respect Dale's law at the level of individual connections, that the posited inhibitory interneurons would introduce additional delays, and that examining how inter-areal synchronization depends on feedback delay is an important direction for future work.

> *Another weakness is the use of the term "predictions" to indicate features of the model dynamics that are purely described in the context of the model parameters. Although the model's response properties may certainly lead to predictions, I think the term requires a better contextualization in terms of neurophysiology and experimental neuroscience. The Discussion draws very interesting and valuable bridges between neuron morphology, interneuron types, and model parameters. But it seems it's left to the reader to backtrack and figure out which biological mechanisms or experimental manipulations should correspond to changes in input or feedback gain, and how these should be distinguished from possible changes in feedforward gain.*

We thank the reviewer for this comment. Each prediction is now grounded in a candidate experimental manipulation and a measurable readout in a new Discussion subsection, "Experimental predictions of the model", which covers each gain parameter, the frequency-resolved subspace, the routing result and normalization, and states explicitly that the proposed manipulations of the gains are speculative and rest on the hypothesized cell-type mapping. The V2 input gain is addressed in our response to the next point.

> *Relatedly, the manuscript places substantial emphasis on modulation of feedback gain, but does not comparably explore modulation of the feedforward gain, β_2, which regulates the V1-to-V2 drive.*
>
> *This seems important because changes in feedforward gain could also influence communication subspace dimensionality and oscillatory dynamics. Therefore, predictions related to top-down feedback modulations should be taken with a grain of salt.*

We thank the reviewer for this comment. We agree, and have added the V2 input gain to every analysis in which feedback and V1 input gain are varied. New SI Fig. S8 varies β_2 over the same range as β_1 (0.5–1.5) and reports, side by side with feedback and V1 input gain, the contrast-response functions of V1 and V2, the V1 power spectra and V1–V2 coherence spectra

at 3% and 50% contrast, and the inter- and within-area prediction performance at 100% contrast. V2 input gain has a signature distinct from both feedback and V1 input gain:

- Increasing β_2 leaves V1 responses unchanged and enhances only V2 responses (Fig. S8c), whereas feedback and V1 input gain enhance both areas.

- At 3% contrast, β_2 produces neither the alpha peak in V1 power nor the beta-band coherence peak that appear at low feedback gain (Fig. S8f,l).

- At 50% contrast, increasing β_2 shifts the gamma peak in power and coherence to higher frequencies and broadens it, like feedback gain, but also reduces its amplitude (Fig. S8i,o).

- Increasing β_2 enhances inter-areal communication, like feedback gain, but leaves within-area communication approximately unchanged (Fig. S8r), whereas feedback gain reduces within-area communication and V1 input gain reduces both.

These results are reported in the Results ("Contrast response functions", "Inter-areal coherence", "Communication subspaces"; the "V1 power spectra" section refers the reader to SI Fig. S8 for the V2 input gain), in the Discussion (opening summary, "Role of feedback" and "Experimental predictions of the model").

> *Last but not least, the model dynamics are split among multiple elements and nonlinear interactions, reaching a level of complexity far higher than the other ORGaNICs formulations present in the literature. The authors derive these dynamics in the supplementary material, as a dynamical system that converges to a fixed-point solution that includes "exact divisive normalization". I wonder, however, if there could be simpler solutions that also produce normalization, either approximate or in a different form than the one proposed by the authors. Note also that the designation of "excitatory neurons" is misleading: despite the presence of two explicitly inhibitory populations, the "excitatory" units also interact with negative effective weights both recurrently and in the inter-areal interactions, thus breaking Dale's law.*

We thank the reviewer for the comment. The complexity is, moreover, not unstructured: as now stated in the Results ("Theory"), each population performs a defined operation — principal cells y sum the gain-modulated input, recurrent and feedback drives, u accumulates the normalization signal, a implements the division, and q provides the sign inversion required by the feedback term in the dynamics of a, and it is through these populations that normalization is implemented dynamically, and simultaneously across the reciprocally coupled areas, rather than imposed as a static input–output relation. Whether a materially simpler circuit could produce normalization, approximately or in a different form, while retaining analytical tractability is an open question that we do not resolve here; we would regard such a circuit as a valuable simplification of the present framework. On terminology: we agree, and have changed it. The y population is now called "principal cells" (or "principal neurons") throughout the main text, figure captions and SI; "excitatory" is reserved for the anatomically excitatory populations in the hypothesised cell-type mapping (Discussion, "Biological plausibility"); the distinction between effective and anatomical weights is stated where the matrices are introduced (Results, "Theory"); and the Limitations state that the principal cells interact through mixed-sign effective weights both recurrently and across areas.

> ***Recommendations for the authors:***
>
> ***Reviewer #1 (Recommendations for the authors):***
>
> *First of all, my sincere apologies for my long delay. It has been quite a hectic period.*

> *If I could suggest one improvement, it would be to add an early conceptual figure and/or minimal toy model that illustrates the core mechanism before introducing the full circuit. For the illustration, I would try to get inspiration from illustrations of the original Semedo et al paper. Such a figure could show how divisive normalization shapes the covariance/cross-power structure, how this affects coherence, and how the resulting eigen spectrum gives rise to a low-dimensional communication subspace. Starting from this stripped-down mechanism would make the novel insight more transparent, after which the more biologically motivated model would be easier to appreciate.*

We thank the reviewer for this suggestion. We agree that the core mechanism should be stated before the full circuit is introduced, and have done so in the text rather than with a new figure. A new paragraph, "The core mechanism in brief", closes the "Theory" section (Results), before any result is presented. It follows the progression the reviewer outlines: the noise-driven fluctuations about the normalized fixed point are characterized by the cross-spectral density matrix (Eq. 6); the power spectra are its diagonal entries, the coherence its normalised off-diagonal entries (Eq. 7), and the communication subspace follows from its frequency-resolved covariance through reduced rank regression, so that power, coherence and the communication subspace are complementary readouts of the same underlying dynamics. How normalization shapes this matrix, by creating a resonance and by suppressing weakly driven neurons, which lowers its rank, is stated in the Discussion ("Unified view of inter-areal communication"), and removing normalization removes all three effects (SI Fig. S4e–h). The full circuit is then introduced as the implementation of this mechanism. We have not added a schematic figure, because the corresponding panels already exist: the with/without-normalization comparison (SI Figs. S4 and S5).

> *Minor comments/suggestions:*
>
> *The authors forgot to cite other work modelling communication subspaces. Citing them could improve impact. Disclaimer: I am among the authors of one of them.*
>
> *https://www.nature.com/articles/s41467-023-42519-5*
>
> *https://www.science.org/doi/full/10.1126/sciadv.adl6120*
>
> *https://www.pnas.org/doi/10.1073/pnas.2404039122*
>
> *https://www.biorxiv.org/content/10.64898/2026.05.18.725829v1.abstract*

We thank the reviewer for these references. We have added all four (Barbosa et al. 2023; Gozel & Doiron 2024; Clark & Beiran 2025; Mastrogiuseppe et al. 2026) in the Introduction, where communication subspaces are introduced.

> *It could improve the impact if the authors could flesh out how they would test 'experimentally testable predictions' (e.g., lines 162, 329).*

We thank the reviewer for this suggestion. We agree, and have revised accordingly. Each prediction is now paired with a candidate manipulation and a readout in a new Discussion subsection, "Experimental predictions of the model".

> *Figure 2 c-d and e-f could make the message stronger if a correlation between model and data is shown: how good is the prediction?*

We thank the reviewer for this suggestion. We agree that a quantitative measure of agreement is needed, and have provided it as a fit rather than a correlation coefficient. New SI Fig. S7 and SI Sec. III fit the model to the experimental power and coherence spectra of Roberts et al. (2013) at all measured contrasts and show the residuals. The Fig. 2 caption and the Results ("V1 power spectra", "Inter-areal coherence") now cite this comparison.

> *Figure 4c is very interesting and deserves more spotlight (could even be tested with available data?)*

We thank the reviewer for this suggestion. We agree, and have given this result more prominence and specified the test. It is stated in the Abstract in its own right. The Methods ("Frequency decomposition of the communication subspace") now end with a paragraph that states how the analysis may be applied to experimental recordings: from the cross-spectral density matrix of simultaneously recorded V1 and V2 populations, its real part at each frequency is the frequency-specific covariance, whose source, target and cross blocks give the communication subspace, the prediction performance and the dimensionality at that frequency in closed form (Eqs. 20 and 22); no band-pass filtering or per-band regression is needed. The resulting test, that the dimensionality should be smallest in the band of peak V1–V2 coherence and that this band should shift to higher frequency with contrast, is stated at the end of that Methods paragraph and in the Discussion ("Experimental predictions of the model"). We have not performed this analysis, which requires the raw simultaneous recordings.

> ***Reviewer #2 (Recommendations for the authors):***
>
> *Major:*
>
> *(1) If possible, try to add at least one quantitative model-data comparison (variance explained) even without fitting if needed, so readers can gauge agreement beyond visual inspection.*

We thank the reviewer for this comment. We agree, and have added a quantitative comparison. New SI Fig. S7 and SI Sec. III fit the model parameters to the experimental power and coherence spectra measurements and report the residuals at every contrast.

> *(2) Show how the gamma-peak predictions change under physiologically realistic time constants (e.g., 10-20 ms), so that readers get an idea.*

We thank the reviewer for this suggestion. The Limitations subsection of the Discussion now explains that if every time constant of the model is multiplied by a common factor k then every spectral feature would be shifted to a frequency k times lower with its shape preserved. The Limitations also now state that when all time constants were released as free parameters in the fit to the experimental spectra (SI Fig. S7) they converged to values between 0.2 and 3ms. The uniform 1ms values are now described throughout as effective population time constants. We decided not to add a figure because it would look identical to Figs. 2 and 3 with a relabelled frequency axis.

> *(3) For the communication-subspace results, report how dimensionality estimates scale with the number of modeled/subsampled neurons, so the comparison to the experimental literature is calibrated.*

We thank the reviewer for this suggestion. We agree, and have added the analysis as new SI Fig. S9. Fig. S9 reports the dimensionality of the inter-areal (V1–V2) and within-area (V1–V1) communication subspaces (i) as a function of the number of modeled principal neurons per area, N = 9, 18, 36, 72 and 144, at the sampling fraction of the main text (30/72 of the neurons as source and as target, i.e. 4–60 neurons; the connectivity kernels and normalization pool are defined on the ring and scale with N, with no parameter re-tuned), and (ii) as a function of the number of sampled source and target neurons, 5–30, at N = 72. Each point is the mean ± SD over 200 random draws of source and target populations at 100% contrast and baseline parameters, with dimensionality defined as the rank at which prediction performance reaches 95% of its full-rank value. The dimensionality grows slowly with both quantities (roughly two-fold over the 16-fold change in N), remains far below the number of sampled

neurons, and the inter-areal subspace is lower-dimensional than the within-area subspace at every population size.

> *Minor:*
>
> *(1) Qualify quantitative-sounding language to reflect that comparisons are qualitative and unfitted.*

We thank the reviewer for this comment. We agree, and have revised accordingly. The unfitted comparisons are now described as qualitative: "successfully replicates key experimental observations" now reads "reproduces the main qualitative features" (Results, "V1 power spectra"), the captured trends and observations are described as qualitative in the power-spectra and coherence sections, and "align well with" now reads "qualitatively consistent with the experimental data" (Results, "Communication subspaces"). The fitted comparison of SI Fig. S7 is cited alongside the power-spectra and coherence comparisons.

> *(2) Clarify in the main text the scope of the "exact normalization" claim*

We thank the reviewer for this comment. We agree, and have revised accordingly. At first mention (Results, "Theory") the main text now states that the implementation is exact for W_{11} = I with $\beta = \gamma = 1$, that for the centre-surround W_{11} used in all reported results the circuit closely approximates normalization, and that the fixed point for the parameters used is obtained numerically by integrating the deterministic dynamics to steady state.

> ***Reviewer #3 (Recommendations for the authors):***
>
> *(1) Abstract: I think the text could be improved in terms of clarity and readability. For example, the (f) for frequency appears only after the $1/f^4$, and similarly for the explanation for "communication". The term "high-coherence" is not properly defined.*

We thank the reviewer for this comment. We agree, and have revised the Abstract. Frequency and its symbol f are introduced together before $1/f^4$ appears; "communication" is defined at first use ("the ability to linearly predict neural activity in one brain area from neural activity in another"); "high-coherence frequencies" is replaced by "the frequencies at which V1–V2 coherence peaks".

> *(2) L50 and L56: The term "a priori constraints" seems to refer to two general features of the model. The model, however, is much more constrained in terms of parameters and architecture. So the statement that "these two constraints alone give rise to a rich set of emergent properties" is not really corroborated by the results.*

We thank the reviewer for this comment. We agree, and have revised the Introduction. It now states the two assumptions (a hierarchy with feedforward, recurrent and feedback connections; recurrently implemented normalization within each area) and then notes that a concrete model additionally requires choices about the connectivity, the normalization pool, and the noise, which are specified in the Methods and whose influence on the results is examined in SI Fig. S4 and the SI section "Robustness with respect to model parameters". The sentence "Remarkably, these two constraints alone give rise to..." now reads "Given these choices, and a single set of parameters, the theory gives rise to...".

> *(3) L69: "competing hypotheses", but they were called "complementary hypotheses" in L34.*

We thank the reviewer for this comment. We agree, and have made the wording consistent. Communication through coherence and communication subspaces are now described, in the Introduction and in the Discussion ("Unified view of inter-areal communication"), as hypotheses that have been treated as alternatives in the literature and that our results show to be complementary.

> *(4) L132: "all neurons perform the same computation"; this comment is a bit too generic and dismissive.*

We thank the reviewer for this comment. We agree, and have rewritten the sentence as a neutral statement where it first occurred (Results, "Theory") and removed the duplicate sentence from the Discussion ("Biological plausibility"). In Wilson–Cowan-type excitatory–inhibitory models the same form of rate equation is applied to each population, with different signs and weights, whereas here the four populations perform different roles with structurally different equations.

> *(5) L134: "unique computations to each cell type" although it is clear that different cell types have different roles in the overall dynamics, it is not clear to me which is the unique computation attributed to each of them.*

We thank the reviewer for this comment. The computation performed by each cell type is now stated at that point in the text (Results, "Theory"): principal cells y sum the gain-modulated input, recurrent and feedback drives (squaring output nonlinearity); excitatory modulator cells u accumulate the normalization signal, a weighted sum of principal-cell responses (square-root output nonlinearity); inhibitory modulator cells a implement the division, i.e. the multiplicative gain control of the recurrent and feedback drives; and inhibitory cells q track principal-cell activity and provide the sign inversion required by the feedback term in the dynamics of a (SI Eqs. S21–S22). Additionally, they all have different activation nonlinearities.

> *(6) L146: "positive and negative weights": mixed signs for feedforward and feedback weights.*

We thank the reviewer for this comment. We agree, and have revised accordingly. The phrase now reads "an embedding matrix of effective weights", followed by a new sentence stating that all weight matrices describe effective connectivity rather than anatomical synapses and that negative weights may be implemented with disynaptic inhibition through interneurons that are not modelled explicitly, with a cross-reference to the Discussion. The same statement accompanies the definition of the matrices in the Methods ("The Model").

> *(7) L160: Again mixed signs for recurrent weights.*

We thank the reviewer for this comment. We agree, see the response above.

> *(8) L189: Normalization is exact, but only for parameters different from those used. I think it should be added there that the fixed point for the parameters in use is found numerically through simulation of the dynamics.*

We thank the reviewer for this suggestion. We agree, and have added the statement in those terms. The passage now states that the implementation is exact only for $W_{11} = I$ with $\beta = \gamma = 1$, that for the centre-surround W_{11} used in all reported results the circuit closely approximates normalization (deviates from Eq. 2 by less than 0.6%), and that the fixed point for these parameters is obtained numerically by integrating the deterministic dynamics to steady state, after which all spectra and subspaces are computed in closed form by linearising about it. The same statement is made in the Methods, in the new subsection "Numerical computation of fixed points".

> *(9) L212: There should be something more explicit to corroborate the "experimentally testable".*

We thank the reviewer for this comment. We agree, and have revised accordingly. Each prediction is now paired with a manipulation and a readout in a new Discussion subsection, "Experimental predictions of the model", which is cited where the predictions are stated in the Introduction and the Discussion.

> *(10) L214: I think it's important to clarify a bit more that even if both areas have input gains, the manipulation is only of the input gain in the upstream area, β_1.*

We thank the reviewer for this comment. We agree, and have clarified this at that point in the text ("with the V2 input gain β_2 held at its baseline value") and in the captions of Fig. 3 and SI Figs. S1–S3, each of which now states that β_2 is held at 1.0. β_2 is varied separately in new SI Fig. S8.

> *(11) Figure 2: Please consider using a better color scale for the power and the coherence lines. The lower three contrasts are literally indistinguishable. Same for the top two contrasts.*

We thank the reviewer for this suggestion. We have retained the palette, for the following reasons. The palette is the "Reds" sequential scheme of matplotlib, derived from the ColorBrewer scheme of the same name (Harrower & Brewer, 2003). The curves encode an ordered variable (contrast), for which a single-hue sequential palette is the recommended choice (Zeileis et al., 2020). This encoding is also colour-vision-deficiency safe: the curves are distinguished by lightness rather than by red-versus-green discrimination, and the ColorBrewer "Reds" scheme is classified as colourblind-safe on the basis of empirical testing with colour-vision-impaired viewers (Olson & Brewer, 1997). The same palette is used in the new SI Figs. S6 and S7 for consistency.

> *(12) L273: Something that confuses me is how to disambiguate between input gain and input intensity, since the theory considers effects around fixed points. This equivalence is also mentioned in the discussion.*

We thank the reviewer for this comment. In the model, the V1 input gain β_1 enters the dynamics only through the product $\beta_1 z_1$ (Eq. 1a), and the input drive z_1 is proportional to stimulus contrast (Methods). Therefore, V1 input gain and stimulus intensity are mathematically equivalent because they enter the dynamics only through their product (β_1 is scalar). This equivalence would not generally hold for spatially or feature-selective gain modulation, where attention can differentially modulate subsets (β_1 is vector) of neurons relative to the normalization pool (Reynolds & Heeger, 2009).

> *(13) L375: There seems to be a lot of emphasis on modulation of the feedback gain, but related to L214, I wonder why there has been no manipulation of the feedforward gain, β_2, which might be equally important in influencing the communication subspace dimensionality and in modulating the oscillatory dynamics. In terms of experimental predictions, the two might be very different, so why consider only one and ignore the other? See also statement at L388. To attribute the effect to top-down feedback, one should prove that bottom-up modulation does not produce the same effect.*

We thank the reviewer for this comment. We agree with the methodological point in the final sentence, and have acted on it. β_2 is now varied across the same analyses as γ_1 and β_1 (new SI Fig. S8; outcomes listed in our response to Weakness 4 above). Bottom-up modulation does not produce the same effect: increasing β_2 enhances inter-areal communication, as feedback gain does, but leaves within-area communication unchanged (feedback gain reduces it), leaves V1 responses unchanged, produces no low-contrast alpha or beta peaks, and reduces rather than increases the amplitude of the gamma peak. The passage at L375 (Results,

"Communication subspaces") now reports that increasing β_2 enhances inter-areal communication while leaving within-area communication approximately unchanged, with a substantially smaller effect than feedback gain (SI Fig. S8r); the readouts that distinguish the two gains are described in the Discussion ("Experimental predictions of the model").

> *(14) Figure 4: See comment for Figure 2.*

As for Fig. 2, we have retained the palette; see our response to item (11). The contrast values are given in the legend, the luminance ordering is monotonic in contrast, and the same palette is used across Figs. 2 and 4 and the SI.

> *(15) L440: I think here it would be more honest to say that the two measures, of coherence and inter-areal communication, are not independent. If two signals are highly correlated at a certain frequency, then the prediction power will also peak for that same frequency. Or would it be possible to have conditions where it would not be the case, showing that the two measures are in fact independent?*

We thank the reviewer for this comment. We agree, and now state the dependence explicitly. The passage at L440 (Results, "Frequency decomposition of communication subspaces"), the Discussion ("Unified view of inter-areal communication") and the Methods ("Interpretation of the results") now state that coherence (Eq. 7) and frequency-resolved prediction performance (Eq. 22) are functionals of the same cross-spectral density matrix, so that the coincidence of their peaks is a mathematical consequence rather than an independent empirical confirmation. The claim retained is that coherence, prediction performance and the reduction in dimensionality are all generated by normalization; removing it eliminates all three (SI Fig. S4e–h).

> *(16) L446: This paragraph mentions dimensionality, but only the following paragraph defines it.*

We thank the reviewer for this comment. We agree, and have moved the definition (the number of source dimensions required to reach the full-rank prediction performance, dictated by the rank of $\mathbf{C}_3^{\mathrm{T}}\mathbf{C}_1^{-1}\mathbf{C}_3$) so that it precedes the first use of "dimensionality" (Results, "Frequency decomposition of communication subspaces"); the sentence that previously followed has been removed.

> *(17) L485: See comment for L375.*

We address this together with item (13) and the corresponding point of the public review (Weakness 4). We have added a statement regarding the V2 input gain at the end of the section "Dynamic modulation of functional connectivity through feedback" (Results), and the conclusion of that section now describes the three-area result as a demonstration that differential feedback strength can control functional connectivity.

> *(18) L575: I find this part a bit too vague. Within the scope of the paper, it is not clear what the network is "processing" and what kind of "balancing" it performs.*

We thank the reviewer for this comment. In this passage, "processing" refers to the computation performed by the principal cells: each cell sums its gain-scaled input, recurrent and feedback drives (Eq. 1a). "Balancing" refers to how these drives are combined: the recurrent gain control adjusts them so that the steady-state responses follow the normalization equation (Eq. 2), exactly or approximately depending on the recurrent weights. This is the mechanistic contrast with predictive coding: predictive coding matches top-down predictions against feedforward signals so as to minimize the prediction error, whereas hierarchical ORGaNICs balances the input, recurrent and feedback drives so as to maintain normalization.

> *(19) L583: Similarly to above, I find these lines very confusing. In the context of this paper, there is no notion of "predictable stimulus", so I find this part very confusing. Also, the model always produces sustained activity by construction, since it has stable steady-state solutions across a wide range of input patterns.*

We thank the reviewer for this comment. The passage refers to a limitation of the predictive coding framework. A constant, continuously presented stimulus is fully predictable, so in predictive coding the forward-propagated activity, which represents the prediction error, should be explained away and decay toward zero; cortical neurons instead maintain sustained responses to constant stimuli, barring adaptation (Burns, Xing & Shapley, 2010). We agree that our model produces sustained responses by construction, because it has a stable fixed point for a wide range of inputs; this behaviour is closer to the observed physiology than the explaining-away account.

> *(20) L601, 602: I don't think that not producing mathematically exact normalization is a significant, critical limitation in a model. We can be pretty confident that neural circuits are not, in fact, producing mathematically exact normalization. Or, to say it more constructively, simpler models that produce approximate normalization might actually be closer to the biological reality.*

We thank the reviewer for this comment. We agree, and have rewritten the passage (Discussion, "Normalization"). It no longer presents approximate normalization as a critical limitation of prior circuit models; it instead states how the present circuit differs: the normalization equation, with an arbitrary normalization pool, is a fixed point of the recurrent dynamics; the dynamics remain stable even with strong recurrent connections; and the fixed point and the linearization about it yield the power spectra, coherence, and communication subspaces in closed form.

> *(21) L653: this line states that feedback is essential for network stability. But the paragraph seems to refer to top-down feedback drive, which should not be necessary for dynamical stability in the model.*

We thank the reviewer for this comment. The line refers to the feedback drive onto the inhibitory modulator cells (a), which contributes to the local normalization signal; it is this feedback-mediated inhibition, not the top-down drive to the principal cells, that is essential for stability. We have clarified the sentence accordingly (Discussion, "Role of feedback").

> *(22) L661: I think this statement is misleading and this part should be stated more cautiously. Although plausible in general, I don't see these results anywhere in the paper. It would require a more functional connotation of the model, which now has a very homogeneous ring structure.*

We thank the reviewer for this comment. We agree, and have stated the claim more cautiously (Discussion, "Role of feedback"): the passage now describes the generative-model regime as a hypothesis ("the network is hypothesized to behave like generative models...") rather than as a property demonstrated by the present results.

> *(23) L715, L719: I feel confused by the contrast between spatially global feature-based attention and spatially-selective attention.*

We thank the reviewer for this comment. The two forms of attention differ in the domain over which neurons are selected for gain modulation. Feature-based attention is spatially global: attending to a feature boosts the neurons selective for that feature across the entire visual field, regardless of retinotopic location (Saenz, Buracas & Boynton, 2002). Spatially selective attention instead boosts only the neurons whose receptive fields overlap the attended location, regardless of their feature preference. Both are gain modulations that can be implemented in the model through the input gain: applied as a vector rather than a scalar,

β_1 raises the gain of a selected subset of neurons relative to the normalization pool (Reynolds & Heeger, 2009), as noted in our response to item (12). On the ring, a β_1 profile over preferred orientation implements feature-selective gain; implementing spatially selective attention in the same way would require adding a spatial (retinotopic) dimension over which β_1 varies, as now noted in the Discussion ("Attention and functional connectivity"). The three-area routing result corresponds to the spatially global case, in which boosting an entire feature-selective area increases its feedback to V1 at all locations.

> *(24) L729, L730: Here it is important to define exogenous and endogenous attention. If they are defined by the pathways on which they act, then the reasoning becomes circular.*

We thank the reviewer for this comment. We agree, and have defined both terms behaviourally at first use, independently of the model (Discussion, "Attention and functional connectivity"): exogenous attention as the fast, involuntary orienting driven by a salient stimulus, and endogenous attention as the slower, voluntary, goal-directed deployment of attention. The mapping onto input gain and feedback gain, respectively, is then stated as a testable hypothesis rather than as a definition, falsifiable through the divergent spectral signatures of the two gains.

> *(25) Figure 7: I suggest changing the color map into a "diverging" color map centered at zero, so that negative and positive values appear visually distinct and zero has a neutral color. For example, the "seismic" or "PuOr" color maps in matplotlib would be appropriate.*

We thank the reviewer for this suggestion. We agree, and have revised Fig. 7 accordingly. It now uses the PuOr colour map, centred at zero with a shared colour bar, so that positive and negative weights in the tuning-curve, recurrent and feedback matrices are visually distinct and zero is neutral.

References cited in this letter

Burns, S.P., Xing, D. & Shapley, R.M. (2010). Comparisons of the dynamics of local field potential and multiunit activity signals in macaque visual cortex. Journal of Neuroscience 30, 13739–13749.

Harrower, M. & Brewer, C.A. (2003). *ColorBrewer.org*: an online tool for selecting colour schemes for maps. The Cartographic Journal 40, 27–37.

Olson, J.M. & Brewer, C.A. (1997). An evaluation of color selections to accommodate map users with color-vision impairments. Annals of the Association of American Geographers 87, 103–134. Polsky, A., Mel, B.W. & Schiller, J. (2004). Computational subunits in thin dendrites of pyramidal cells. Nature Neuroscience 7, 621–627.

Reynolds, J.H. & Heeger, D.J. (2009). The normalization model of attention. Neuron 61, 168–185.

Saenz, M., Buracas, G.T. & Boynton, G.M. (2002). Global effects of feature-based attention in human visual cortex. Nature Neuroscience 5, 631–632.

Silver, R.A. (2010). Neuronal arithmetic. Nature Reviews Neuroscience 11, 474–489.

Zeileis, A., Fisher, J.C., Hornik, K., Ihaka, R., McWhite, C.D., Murrell, P., Stauffer, R. & Wilke, C.O. (2020). colorspace: a toolbox for manipulating and assessing colors and palettes. Journal of Statistical Software 96, 1–49.

Srinivasan, M.V. & Bernard, G.D. (1976). A proposed mechanism for multiplication of neural signals. Biological Cybernetics 21, 227–236.

https://doi.org/10.7554/eLife.111297.2.sa0