# One Nose but Two Nostrils: Learning to Align with Sparse Connections between Two Olfactory Cortices

Bo Liu [1,2] Shanshan Qin [3,*] Venkatesh N. Murthy [1,2,†] and Yuhai Tu [4,‡]

[1] *Center for Brain Science and Department of Molecular and Cellular Biology, Harvard University, Cambridge, Massachusetts 02138, USA*
[2] *Kempner Institute for the Study of Natural and Artificial Intelligence, Harvard University, Cambridge, Massachusetts 02138, USA*
[3] *John A. Paulson School of Engineering and Applied Sciences, Harvard University, Cambridge, Massachusetts 02138, USA*
[4] *IBM Thomas J. Watson Research Center, Yorktown Heights, New York 10598, USA*

(Received 9 May 2024; accepted 22 October 2024; published 2 December 2024)

The integration of neural representations in the two brain hemispheres has been studied extensively in vision, audition, and somatosensation, but less so in olfaction. Recent experiments have revealed that odor responses in cortical neurons driven by separate stimulation of the two nostrils are highly correlated. This bilateral alignment points to structured, nonrandom interhemispheric connections, but how this structure leads to alignment is unclear. Here, we hypothesized that continuous exposure to environmental odors shapes these projections and modeled it as online learning with a local Hebbian rule. We found that Hebbian learning with sparse connections achieves bilateral alignment, exhibiting a linear tradeoff between speed and accuracy (lower learning rate leads to higher alignment level but slower convergence). Furthermore, we identified an inverse scaling relationship between the number of cortical neurons and the interhemispheric projection density required for desired alignment accuracy. Thus, more cortical neurons allow sparser interhemispheric projections, which was explained analytically. We next compared the alignment performance of local Hebbian rule and the global stochastic gradient descent (SGD) learning rule used in artificial neural networks. We found that although SGD leads to the same alignment accuracy with a slightly reduced sparsity, the same inverse scaling relation holds. Our analysis showed that their similar performance originates from the fact that the update vectors of the two learning rules align with each other throughout the entire learning process. Our work suggests that a biologically plausible mechanism with sparse connections suffices for correlated bilateral responses. The quantitative comparison between the local Hebbian rule and the global SGD rule may inspire efficient sparse local learning algorithms for more complex problems.

DOI: 10.1103/PRXLife.2.043016

## I. INTRODUCTION

Animals with bilateral symmetry must integrate information from the two sides of the brain to form a coherent picture of the environment. This problem has been studied in vision and audition, where similarities and asymmetries in information sampled through the two sides can offer interpretable computational clues [1–4]. But it is less clear how information from two sensors is integrated in olfaction [5–9].

In many species including rodents, the two nares are separated by a septum that prevents inhaled odors from transferring between nostrils. Since the two nostrils may sample independent pockets of air [10–12], bilateral information can be used, in principle, to discover features of the odor stimulus not readily available to a single nostril. By comparing odor concentration across the two nostrils, foraging animals may locate odor sources under conditions where smooth gradients exist [8,13–16]. Such asymmetric information processing from the two nostrils has also been observed in humans, who can sense the direction of an odor source presented close to their nose [6,17]. In contrast to computations emphasizing the differences between the two sides, many behaviors require making common inferences that are independent of the stimulated nostril. For instance, animals must be able to identify relevant smells of food or a predator, regardless of the odor source's location relative to their nose and independent of any asymmetries in air flow through the nostrils [7,18].

These behaviors using two nostrils require neural computations that rely on interhemispheric communication, which occurs via axons crossing through the anterior commissure [8,9,19]. Two brain regions, anterior olfactory nucleus (AON) and anterior piriform cortex (APC), are strong candidates to integrate information from the two nostrils [9,10,20–24]—we refer to these regions collectively as the olfactory cortex (OC). The olfactory bulb (OB), the first stage of circuit processing in the olfactory system, sends projections exclusively to the ipsilateral OC ("ipsi-lateral" means "on the

*Present address: Center for Computational Neuroscience, Flatiron Institute, New York, New York 10010, USA.
†Contact author: vnmurthy@fas.harvard.edu
‡Contact author: yuhai@us.ibm.com

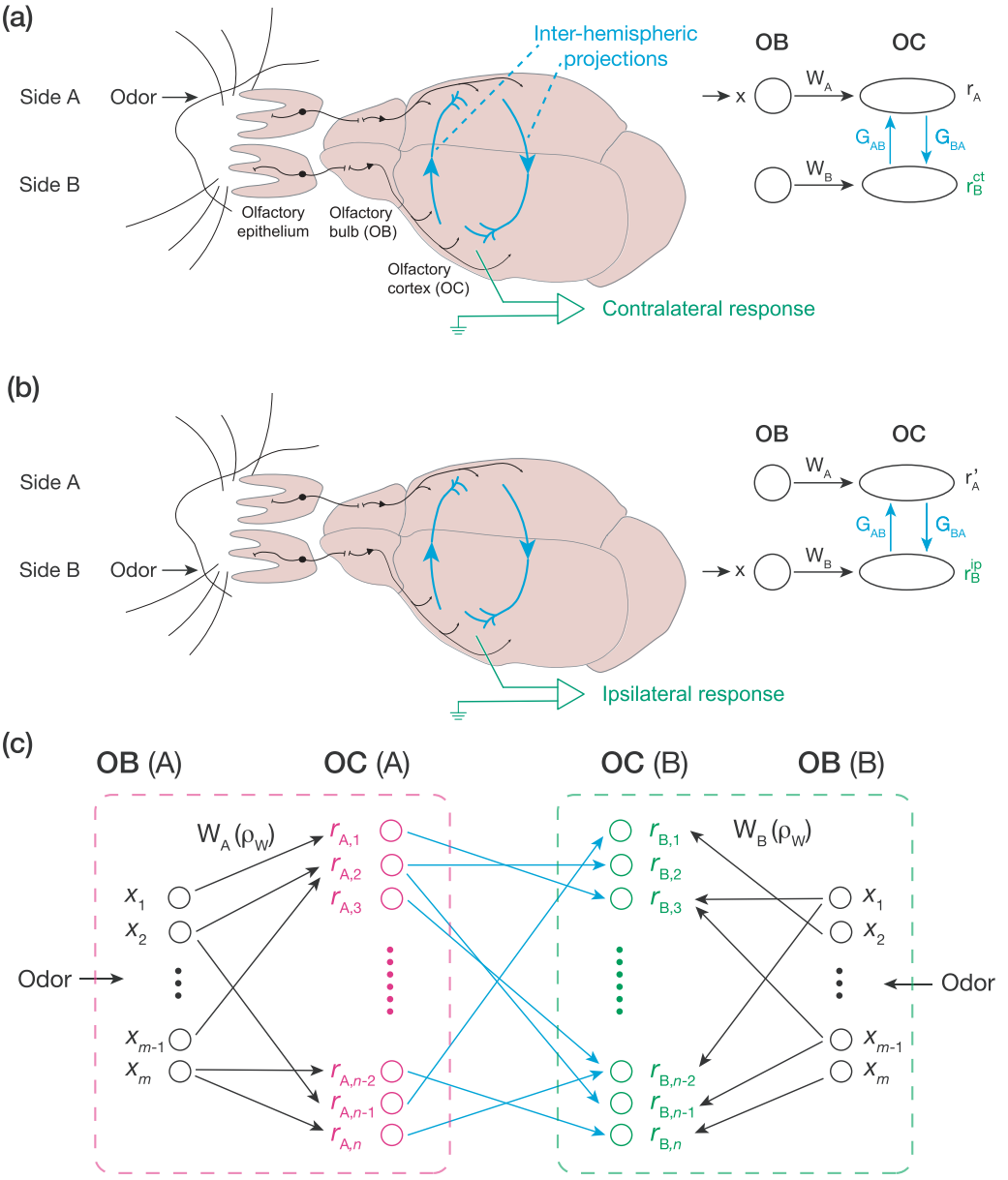


FIG. 1. Background and model setup. (a) A sketch of the early mouse olfactory system and the experimental setup to measure the contralateral response (left), and the corresponding neural circuit (right). Odor is delivered to the nostril on side A, and the contralateral response is measured by the electrode placed in the side-B olfactory cortex. (b) A sketch of the experimental setup to measure the ipsilateral response (left), and the corresponding neural circuit (right). (c) Neural network of the bilateral alignment circuit during training. The pink box represents Side-A hemisphere, and the green one is for Side-B. Training odors are delivered to both nostrils. Due to the two random and different OB-to-OC projections ($\mathbf{W}_A$ and $\mathbf{W}_B$), the cortical representations $\mathbf{r}_A$, $\mathbf{r}_B$ are also different. The blue arrows ($\mathbf{G}_{BA}$) illustrate the inter-hemispheric projections that Side-B OC receives from Side-A OC, with density $\rho_G$. $\mathbf{G}_{AB}$ is not shown in this network. Adapted from [36].

same side," versus "contra-lateral"). Neurons in the OC, however, send projections to the opposite hemisphere through the anterior commissure and make functional synapses [20,24–27].

Single OC neurons can respond to odors presented separately to ipsilateral nostril or contralateral nostril. Although direct projections from the OB account for ipsilateral responses, interhemispheric projections are the likely substrate for the responses of OC neurons to odors presented to the contralateral nostril [9,10,21,22,28] [Fig. 1(a)]. However, given the apparently random and different OB-to-OC projections on two sides [29–33], it is unclear whether and how the neuronal responses to ipsi- and contralateral stimuli are correlated in the OC, and if so, how similar they are. On the one hand, systematically different responses can be used to compute the direction from which an odor stimulus arises in natural situations [21]. On the other hand, useful computations may not be possible if the responses are entirely uncorrelated. Local chemical gradients are not informative for animals in many natural situations, particularly in turbulent environments with highly intermittent odor detections [34,35]. In these conditions an animal may benefit from appropriate averaging of

noisy information from the two nostrils, which requires the two sides to have some degree of correlated responses.

Recently, we used a panel of diverse odors to probe the structure of odor responses in individual mouse olfactory cortical neurons to ipsilaterally and contralaterally presented stimuli [36]. We found that individual neurons responded to odors presented to the contralateral nostril in a similar way to the same odors presented ipsilaterally [36]. This similar tuning, which we refer to as bilateral alignment, was significantly better than chance. Therefore, matching responses to odors presented separately to the two nostrils must arise from coordinated connectivity in the interhemispheric projections, which must somehow be aligned with the ipsilateral projections. To achieve such coordinated connectivity thus bilateral alignment, an attractive and plausible hypothesis is that Hebbian or correlation-based synaptic plasticity shapes synapses formed by interhemispheric axons [Fig. 1(a)].

In the present work we hypothesized that the continuous exposure of both nostrils to environmental odors shapes the functional properties of interhemispheric projections and tested whether online learning with local Hebb's rule is sufficient to account for matched responses found experimentally. Furthermore, mammals have millions of olfactory cortical neurons [37], while the typical number of synapses onto one neuron is $10^3 - 10^4$ [38,39]; thus, a cortical neuron in one hemisphere is unlikely to connect with all the neurons in the other hemisphere [24,40], leading to the questions of whether sparse interhemispheric connectivity could achieve bilateral alignment, and if so, how sparse or dense the connections need to be. Finally, within the context of bilateral alignment, we studied the differences and similarities between the local Hebb's rule and gradient-based global learning rules, particularly the stochastic gradient descent (SGD) rule (algorithm) that is prevalent in training artificial neural networks.

The paper is organized as follows: First, we define the problem by describing the online Hebbian learning process in a simplified network model of the olfactory cortex and the testing criterion for bilateral alignment. We then present three main findings from our study: (1) Online Hebbian learning can successfully achieve bilateral alignment with highly sparse interhemispheric projections, and the learning rate modulates the tradeoff between alignment accuracy and convergence speed. (2) For a given degree of alignment, the required density of interhemispheric projections scales inversely with the number of cortical neurons, meaning that more cortical neurons allow sparser connections. (3) The SGD learning rule achieves the same alignment performance with modestly sparser connectivity but exhibits the same scaling relation. Furthermore, we showed that the similar performance for local Hebbian rule and global SGD rule is due to the fact that the update vectors of the two learning rules are partially aligned during the learning process.

## II. MODEL SETUP AND TRAINING PROCESS

To study whether online Hebbian learning can lead to bilateral alignment, we use a minimum learning model as illustrated in Figs. 1(a)–1(c), where (a) and (b) illustrate the circuits for contra- and ipsilateral responses, respectively. As shown by the neural network in Fig. 1(c), odor elicited the same OB responses in two sides, denoted as $\mathbf{x} \in \mathbb{R}^m$ (anatomically, $m$ stands for the number of OB glumeruli [41,42], and for simplicity, we refer to the responses of output neurons of the glumeruli as "OB responses"). The projections from OBs to OCs ($\mathbf{W}_\mathrm{A}$ and $\mathbf{W}_\mathrm{B}$) are modeled to be largely random. This leads to different cortical odor representations in two hemispheres $\mathbf{r}_\mathrm{A}, \mathbf{r}_\mathrm{B} \in \mathbb{R}^n$, which are also modulated by the interhemispheric interactions ($\mathbf{G}_\mathrm{AB}$ and $\mathbf{G}_\mathrm{BA}$).

For simplicity, the OB response $\mathbf{x}$ to an odor is drawn from Gaussian distribution with $\mathbf{0}$ mean: $\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma})$ (negative firing rates indicate suppressed activities of OB output neurons, compared to their baseline firing rates). Here $\mathbf{\Sigma}$ is the input covariance matrix, and for analytical tractability, we take $\mathbf{\Sigma} = \gamma^2 \mathbf{I}_m$ (independent and identically distributed or IID Gaussian distribution) with $\gamma$ the input strength. To characterize the sparse OB-to-OC projections [33,43], we denote the density of $\mathbf{W}_\mathrm{A}$ and $\mathbf{W}_\mathrm{B}$ as $\rho_\mathrm{W}$ (the fraction of their nonzero elements). For analytical tractability, we assume that the nonzero elements of $\mathbf{W}_\mathrm{A}$ and $\mathbf{W}_\mathrm{B}$ are also IID Gaussian variables, i.e., $\mathbf{W}_{ij} \sim \mathcal{N}(0, 1), \forall \mathbf{W}_{ij} \neq 0$, which are sampled at the beginning and fixed after that. In contrast, the interhemispheric projection matrix $\mathbf{G}_\mathrm{BA}$ will be updated during learning [the blue arrows in Figs. 1(a)–1(c)], whose connectivity density is $\rho_\mathrm{G}$. We initialize the nonzero entries of $\mathbf{G}_\mathrm{BA}$ with IID Gaussian distribution $\mathbf{G}_{\mathrm{BA},ij} \sim \mathcal{N}(0, \sigma_0^2), \forall \mathbf{G}_{\mathrm{BA},ij} \neq 0$, with a small variance $\sigma_0^2$ to avoid divergence of the dynamics (the same initialization distribution is used for $\mathbf{G}_\mathrm{AB}$). Note that the choice of initialization $\sigma_0^2$ does not affect the steady-state solution of Hebbian learning dynamics.

During learning, each step contains three parts [Figs. S1(b)–S1(c)]: firstly, one training odor is "delivered" to two nostrils [Figs. 1(c) and S1(a)]. The two cortical odor representations, i.e., the firing rate vectors $\mathbf{r}_\mathrm{A}$ and $\mathbf{r}_\mathrm{B}$, are then determined by the steady-state solution of the following neural dynamics:

$$\begin{aligned} \tau \frac{d\mathbf{r}_\mathrm{A}}{dt} &= -\mathbf{r}_\mathrm{A} + \tanh(\mathbf{W}_\mathrm{A}\mathbf{x} + \mathbf{G}_\mathrm{AB}\mathbf{r}_\mathrm{B}), \\ \tau \frac{d\mathbf{r}_\mathrm{B}}{dt} &= -\mathbf{r}_\mathrm{B} + \tanh(\mathbf{W}_\mathrm{B}\mathbf{x} + \mathbf{G}_\mathrm{BA}\mathbf{r}_\mathrm{A}), \end{aligned} \tag{1}$$

where $\tau$ is the single-neuron integration time constant ($\tau = 10\,\mathrm{ms}$ is used in this study), and we adopted the hyperbolic tangent function tanh() as the activation function.

Secondly, assuming a timescale separation between the neural dynamics and learning (synaptic update), we update $\mathbf{G}_\mathrm{BA}$ by a discrete Hebb's rule (similar for $\mathbf{G}_\mathrm{AB}$):

$$\Delta \mathbf{G}_\mathrm{BA} = \eta\left(\mathbf{r}_\mathrm{B}\mathbf{r}_\mathrm{A}^\top - \beta \mathbf{G}_\mathrm{BA}\right) \odot \hat{\mathbf{G}}_\mathrm{BA}, \tag{2}$$

with $\eta$ the learning rate and $\beta$ the weight decay constant. Here, $\odot$ represents element-wise product; $\hat{\mathbf{G}}_\mathrm{BA}$ is a fixed sparsity mask matrix with binary entries (0 or 1, the probability of $\hat{\mathrm{G}}_{\mathrm{BA},ij} = 1$ is $\rho_\mathrm{G}$) to enforce the sparsity of interhemispheric connections (note that $\hat{\mathbf{G}}_\mathrm{BA}$ is also applied to the initial weights). This Hebb's rule is local, because the right-hand side of Eq. (2) involves only the presynaptic and the postsynaptic neurons.

Finally, with the Hebbian-updated $\mathbf{G}_\mathrm{BA}$ and $\mathbf{G}_\mathrm{AB}$, following experiments, we simulate the contralateral response to this single training odor [$\mathbf{r}_\mathrm{B}^\mathrm{ct}$, Fig. 1(a)], as well as the ipsilateral

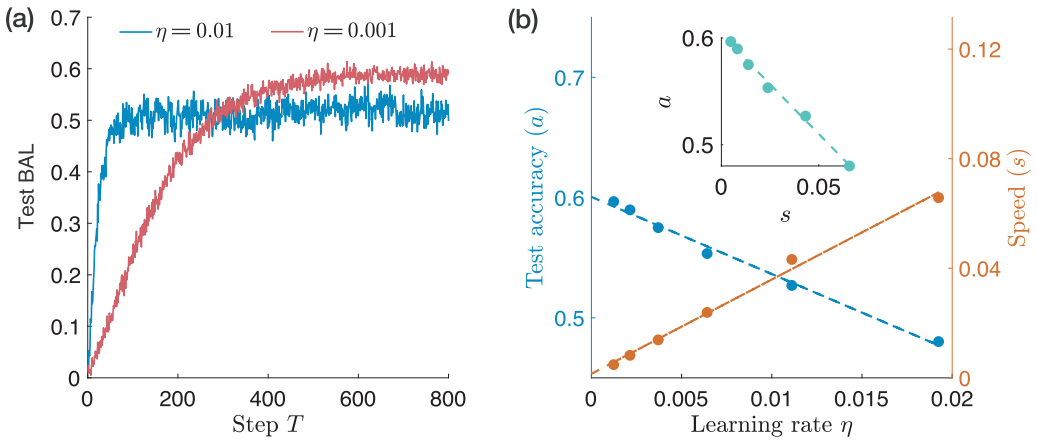


FIG. 2. Hebbian learning can align bilateral responses using sparse connections, with learning rate $\eta$ controlling the linear speed-accuracy tradeoff. (a) Two learning curves with $\eta = 0.01$ s$^2$ (blue) and $\eta = 0.001$ s$^2$ (red), given $(m, n, \rho_W, \rho_G) = (20, 500, 0.1, 0.05)$. (b) Test accuracy ($a$, left $y$-axis) and convergence speed ($s$, right $y$-axis) both depend on $\eta$ linearly but with opposite signs, resulting in a linear tradeoff (the green line of the inset). Here each dot is from numerical results, and the dashed lines are from linear fitting.

response $\mathbf{r}_B^{ip}$ [Fig. 1(b)]. Equations for $\mathbf{r}_B^{ct}$ and $\mathbf{r}_B^{ip}$ can be seen in the Supplemental Material (SM) Sec. I A [44]. To characterize the performance of online Hebbian learning at step $T$, we randomly generate $K$ test odors from the same distribution and then simulate their ipsi- and contralateral responses. We define the test bilateral alignment level (BAL) as the mean cosine similarity $a(T) := \langle \cos(\mathbf{r}_B^{ip}, \mathbf{r}_B^{ct}) \rangle_K$, which characterizes the alignment performance at step $T$.

## III. RESULTS

### A. Dynamics of learning for bilateral alignment: Learning rate modulates the speed-accuracy tradeoff

To study whether sparse interhemispheric projections are sufficient for bilateral alignment, we started with $(m, n, \rho_W, \rho_G) = (20, 500, 0.1, 0.05)$, $N = 1000$ steps, and a learning rate of $\eta = 0.01$ s$^2$. Other parameters are listed in SM Table I [44]. As shown by the blue curve in Fig. 2(a), the test BAL converged to 0.51 within 200 steps, meaning that Hebb's rule can achieve bilateral alignment using sparse connections.

We then adopted a smaller learning rate $\eta = 0.001$ s$^2$, which led to higher test BAL but slower convergence [the red curve in Fig. 2(a)]. Furthermore, we varied $\eta$ with other parameters fixed and found that test accuracy decreases linearly with $\eta$ [the blue dots in Fig. 2(b)]. In contrast, the convergence speed increases linearly with $\eta$ [the red dots in Fig. 2(b), see SM Sec. VIII [44] for the methods to numerically estimate the convergence speed and Fig. S2 for one example]. As a result, test accuracy and convergence speed exhibit a linear tradeoff [the inset green line in Fig. 2(b)].

To understand why Hebb's rule leads to accurate bilateral alignment, we considered a scenario with small input strength $\gamma$ and large decay rate $\beta$, where Eq. (1) can be linearized due to the small magnitudes of $\mathbf{x}$, $\mathbf{G}_{AB}$, and $\mathbf{G}_{BA}$ (see SM Sec. III [44] for details):

$$\tau \frac{d\mathbf{r}_A}{dt} = -\mathbf{r}_A + \mathbf{W}_A\mathbf{x} + \mathbf{G}_{AB}\mathbf{r}_B,$$
$$\tau \frac{d\mathbf{r}_B}{dt} = -\mathbf{r}_B + \mathbf{W}_B\mathbf{x} + \mathbf{G}_{BA}\mathbf{r}_A, \quad (3)$$

which has an exact steady-state solution (SM Sec. I B [44]). Furthermore, with small $\mathbf{G}_{AB}$ and $\mathbf{G}_{BA}$, the steady-state solutions of $\mathbf{r}_A$ and $\mathbf{r}_B$ as well as the ipsi- and contralateral responses become (SM Sec. I B [44])

$$\mathbf{r}_A \approx \mathbf{W}_A\mathbf{x}, \ \mathbf{r}_B \approx \mathbf{W}_B\mathbf{x},$$
$$\mathbf{r}_B^{ip} \approx \mathbf{W}_B\mathbf{x}, \ \mathbf{r}_B^{ct} \approx \mathbf{G}_{BA}\mathbf{W}_A\mathbf{x}. \quad (4)$$

Then Hebb's rule Eq. (2) becomes

$$\Delta\mathbf{G}_{BA} \approx \eta\big(\mathbf{W}_B\mathbf{x}\mathbf{x}^\top\mathbf{W}_A^\top \odot \hat{\mathbf{G}}_{BA} - \beta\mathbf{G}_{BA}\big). \quad (5)$$

In the continuum time limit, the above learning rule can be described by the following stochastic differential equation (SDE):

$$\frac{d\mathbf{G}_{BA}}{dt} = \eta\beta\big(\mathbf{W}_B\mathbf{x}\mathbf{x}^\top\mathbf{W}_A^\top \odot \hat{\mathbf{G}}_{BA} - \mathbf{G}_{BA}\big), \quad (6)$$

where we have made the scaling transformation $\beta\mathbf{G}_{BA} \to \mathbf{G}_{BA}$, which does not affect the cosine similarity. In steady state where $\langle d\mathbf{G}_{BA}/dt \rangle = \mathbf{0}$ ( $\langle\rangle$ means average over time or samples), we have the Hebbian solution

$$\mathbf{G}^\dagger := \mathbf{W}_B\langle\mathbf{x}\mathbf{x}^\top\rangle\mathbf{W}_A^\top \odot \hat{\mathbf{G}}_{BA} = \mathbf{W}_B\mathbf{\Sigma}\mathbf{W}_A^\top \odot \hat{\mathbf{G}}_{BA}, \quad (7)$$

indicating that the Hebb's rule learned two feedforward projection matrices and the input correlation matrix to achieve bilateral alignment. The above analytical solution was verified by numerical simulations (Fig. S3).

To explain the speed-accuracy tradeoff, we vectorize $\mathbf{G}_{BA}$ and write $\mathbf{g}_{BA}(T) := \mathrm{vec}[\mathbf{G}_{BA}(T)] = \mathbf{g}^\dagger + \delta\mathbf{g}(T)$, with $\mathbf{g}^\dagger$ the vectorized Hebbian solution and $\delta\mathbf{g}$ the variation around it. For convenience, we drop the subscript "BA" in $\mathbf{g}_{BA}$ and rewrite the continuum SDE for Hebbian learning [Eq. (6)] as

$$\frac{d\mathbf{g}}{dt} = \eta\beta(\mathbf{g}^\dagger - \mathbf{g}) + \eta\beta\boldsymbol{\xi}, \quad (8)$$

which is just the well-known Langevin equation for an Ornstein-Uhlenbeck process [45] that can be solved analytically. Here, the term $\boldsymbol{\xi} = \mathrm{vec}[\mathbf{W}_B(\mathbf{x}\mathbf{x}^\top - \mathbf{\Sigma})^\top\mathbf{W}_A^\top \odot \hat{\mathbf{G}}_{BA}]$ represents the white noise that originates from input fluctuations (random training odors). Denote its covariance matrix

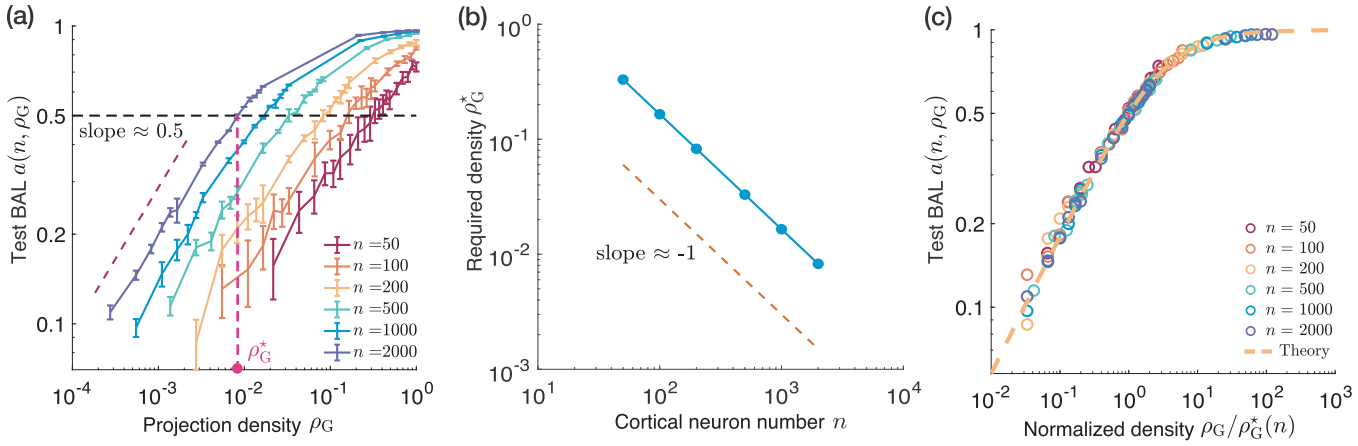


FIG. 3. Numerical and theoretical results of test BAL, and the inverse scaling between $\rho_G^\star$ and $n$. (a) Test BAL increases with $\rho_G$ in a power-law fashion with an exponent around 0.5 for small $\rho_G$. The required density $\rho_G^\star$ to achieve an alignment of 0.5 are determined by the black dashed line, e.g., the purple $\rho_G^\star \approx 8.3‰$ for $n = 2000$. Colors indicate different values of $n$ with fixed $m = 20$, $\rho_W = 0.1$. (b) $\rho_G^\star$ scale inversely with $n$. (c) All the data points in panel A collapse onto a single curve with normalized $\rho_G$ and match well with theoretical prediction (the orange line).

as $\mathbf{cov}(\boldsymbol{\xi}) := 2\mathbf{D}$, where $\mathbf{D}$ is determined by $\mathbf{W}_A$, $\mathbf{W}_B$, $\hat{\mathbf{G}}_{BA}$, $\boldsymbol{\Sigma}$ but independent of $\eta\beta$. It is clear from Eq. (8) that the relaxation timescale to reach the Hebbian solution $\mathbf{g}^\dagger$ is $\tau \sim \frac{1}{\eta\beta}$, which explains the linear dependence of the learning or convergence speed ($s = 1/\tau \sim \eta\beta$) as observed in Fig. 2(b).

By solving Eq. (8), we can show that in steady state ($T \to \infty$), $\mathbf{g}$ has a Gaussian distribution with the Hebbian solution $\mathbf{g}^\dagger$ as its mean and a covariance matrix given by

$$\lim_{T\to\infty} \mathbf{cov}[\delta\mathbf{g}(T)] = \eta\beta\mathbf{D}. \tag{9}$$

To determine the test BAL, we Taylor-expanded $a(\mathbf{g})$ to the second order around its mean $a^\dagger = a(\mathbf{g}^\dagger)$ at the Hebbian solution $\mathbf{g}^\dagger$:

$$a(\mathbf{g}) = a^\dagger + \left(\frac{\partial a}{\partial \mathbf{g}}|_{\mathbf{g}^\dagger}\right)^\top \delta\mathbf{g} + \frac{1}{2}\delta\mathbf{g}^\top\mathbf{H}^\dagger\delta\mathbf{g}, \tag{10}$$

where $\mathbf{H}^\dagger := \mathbf{H}(\mathbf{g}^\dagger)$ is the Hessian matrix evaluated at $\mathbf{g}^\dagger$. Averaging over the Gaussian distribution of $\delta\mathbf{g}$, the stationary test BAL is

$$a(\eta,\beta) = a^\dagger + \frac{1}{2}\langle\delta\mathbf{g}^\top\mathbf{H}^\dagger\delta\mathbf{g}\rangle_{\delta\mathbf{g}} = a^\dagger + \frac{\eta\beta}{2}\mathrm{Tr}(\mathbf{D}\mathbf{H}^\dagger). \tag{11}$$

Both numerical and analytical results showed that the Hessian matrix $\mathbf{H}^\dagger$ is negative definite: concretely, in SM Sec. II C-IID [44], we showed that $a(\mathbf{g})$ is concave around $\mathbf{g}^\dagger$, which is verified numerically in Fig. S4. Furthermore, as a covariance matrix, $\mathbf{D}$ is positive definite. As a result, $\mathrm{Tr}(\mathbf{D}\mathbf{H}^\dagger) < 0$ (see SM Sec. II E [44] for the proof). Therefore, the test BAL decreases with $\eta\beta$ linearly, in agreement with Fig. 2(b).

### B. Larger networks require sparser connections to achieve bilateral alignment: An inverse scaling relation

In the previous section, we demonstrated that the system can achieve bilateral alignment using relatively sparse interhemispheric connections (5% connectivity) in a relatively small network ($n = 500$ neurons). However, in realistic biological systems such as the piriform cortex, there are millions of cortical neurons [37,46], and the interhemispheric projections in mammals might be much sparser than the 5% connectivity (as suggested by the experimental data [24] and the typical $10^3 - 10^4$ synapses onto one neuron [38,39]). Therefore, in this section, the critical question we aim to address is whether bilateral alignment can be achieved in a larger network with sparser interhemispheric connections (smaller $\rho_G$ and larger $n$), or equivalently, how the required sparsity to attain a given alignment performance depends on the network size.

To investigate this question, we varied the number of cortical neurons $n$ and studied its relationship with the required interhemispheric projection density $\rho_G^\star$ to achieve a desired level of test BAL. In this section, we used a small learning rate $\eta = 0.001$ s$^2$ to eliminate the effect of learning rate, i.e., $a \approx a^\dagger$, which depends only on the properties of the network ($\rho_G$, $\rho_W$, $m$, and $n$). For fixed $m$ and $\rho_W$, we varied $\rho_G$ for $n$ ranging from 50 to 2000 and plotted the ensemble average of the test BAL versus $\rho_G$ in Fig. 3(a) (here the ensemble means different samples of the matrices $\mathbf{W}_A$, $\mathbf{W}_B$, $\hat{\mathbf{G}}_{AB}$, $\hat{\mathbf{G}}_{BA}$, given $m$, $n$, $\rho_W$, $\rho_G$). We found that $a(\rho_G)$ has a power-law dependence on $\rho_G$, with an exponent ≈0.5 (the left red dashed line) for small $\rho_G$ before it saturates to near perfect alignment $a \sim 1$. We define a critical $\rho_G^\star(n)$ as the interhemispheric projection density needed to reach a given (high) alignment level $a^\star = 0.5$ (the black horizontal line), i.e., $a(\rho_G^\star(n)) = a^\star$. Our simulation results revealed an inverse scaling relationship between $\rho_G^\star$ and $n$ [Fig. 3(b)], i.e., $\rho_G^\star(n) \sim n^{-1}$, which suggests that with more cortical neurons, the interhemispheric projections can be even sparser. Remarkably, when we rescaled $\rho_G$ by $\rho_G^\star(n)$, i.e., $\frac{\rho_G}{\rho_G^\star(n)} \propto n\rho_G$, all data points in Fig. 3(a) for different values of $n$ collapsed onto one single curve for the full range of sparsity [Fig. 3(c)].

To understand the inverse scaling between $\rho_G^\star$ and $n$ analytically, it is useful to study the alignment between the two vectorized weight matrices $\mathbf{W}_B$ and $\mathbf{G}^\dagger\mathbf{W}_A$, defined as

$$c := \cos(\mathbf{W}_B, \mathbf{G}^\dagger\mathbf{W}_A). \tag{12}$$

Note that we have omitted the notation of vectorization, a convention adopted throughout this paper. It is intuitively clear that $c$ is highly correlated with the bilateral alignment level $a$, e.g., perfect alignment ($a = 1$) is achieved when $c = 1$ (or equivalently, $\mathbf{G}^\dagger \mathbf{W}_\mathrm{A} \propto \mathbf{W}_\mathrm{B}$), since $\mathbf{r}_\mathrm{B}^\mathrm{ct} = \mathbf{G}^\dagger \mathbf{W}_\mathrm{A} \mathbf{x} \propto \mathbf{W}_\mathrm{B} \mathbf{x} = \mathbf{r}_\mathrm{B}^\mathrm{ip}, \forall \mathbf{x}$. Indeed, both simulations and derivations demonstrated that for IID Gaussian input, $a \approx c$ holds true in general (see Fig. S5(a) and SM Sec. II A [44] for details). Using the Hebbian solution Eq. (7) and averaging over independent elements of $\mathbf{W}_\mathrm{A}, \mathbf{W}_\mathrm{B}, \hat{\mathbf{G}}_\mathrm{BA}$ with the law of large numbers, we derived an analytical expression of $c$, which leads to an analytical expression for the BAL at the Hebbian solution $a^\dagger$ (see SM Sec. II B [44] for details):

$$a^\dagger \approx c \approx \left( \frac{n\rho_\mathrm{G}}{m + 3/\rho_\mathrm{W} + n\rho_\mathrm{G}} \right)^{1/2}. \tag{13}$$

The above approximation of $a^\dagger$ agrees well with the numerical results for a wide range of parameters, as shown in Fig. 3(c) [see also Figs. S5(b)–S5(d), where we varied $\rho_\mathrm{W}, \rho_\mathrm{G}$ for given $m = 20, n = 50$].

From Eq. (13), we can see that the required density $\rho_\mathrm{G}^\star$ to achieve a given level alignment $a^\star$ satisfies

$$n\rho_\mathrm{G}^\star = \frac{a^{\star 2}(m + 3/\rho_\mathrm{W})}{1 - a^{\star 2}} = \text{const.} \tag{14}$$

Therefore, $\rho_\mathrm{G}^\star$ scales inversely with $n$. Note that Eq. (13) also explains the power law $a \propto \rho_\mathrm{G}^{0.5}$ in the sparse regime: when $n\rho_\mathrm{G} \ll m + 3/\rho_\mathrm{W}$, $a \approx \sqrt{\frac{n\rho_\mathrm{G}}{m+3/\rho_\mathrm{W}}} \propto \rho_\mathrm{G}^{0.5}$.

The biological interpretation of this inverse scaling is interesting to point out. $n\rho_\mathrm{G}$ in Eq. (13) represents the average number of projections that one cortical neuron receives from the contralateral cortical neurons. Thus, our results show that a good alignment performance (e.g., $a = 0.5$) can be achieved when the average number of connections per neuron is larger than a threshold that is independent of the network size. Therefore, to maintain the same alignment performance, the total number of required interhemispheric connections scales linearly only with $n$ instead of $n^2$, much sparser than all-to-all connections.

As an application of Eq. (13), we estimated the required density of the interhemispheric projections, using biologically realistic parameters [37,43,46,47]: with $m = 3700, n = 5 \times 10^5, \rho_\mathrm{W} = 0.1$, and the required BAL $a^\star = 0.31$, the corresponding $\rho_\mathrm{G}^\star \approx 7.93 \times 10^{-4}$ and $n\rho_\mathrm{G}^\star \approx 397$, meaning ~400 interhemispheric projection synapses per neuron, which is reasonable given the $10^3 - 10^4$ synapses per neuron typically [38,39]. Note that the above estimation is insensitive to the exact value of $\rho_\mathrm{W}$ provided that $m \gg 3/\rho_\mathrm{W}$. The exact sparsity of interhemispheric connections has not been measured: given the many components of OC such as AON and APC, the optogenetic experiments in Ref. [24] showed that only the interhemispheric projections from AON to contralateral AON are sparser than those from AON to ipsilateral APC. Therefore, this rough estimation $n\rho_\mathrm{G}^\star \approx 400$ may inspire future experiments to get more quantitative measurement.

Finally, we confirmed that the fine structure of OC does not change our results: with several subregions and diverse neuron types in the olfactory cortex [24,48], it is likely that only a subpopulation of cortical neurons project contralaterally. While the exact fraction is unknown, we verified that this fraction does not alter the results. We denoted this fraction as $f$ and assumed that these neurons project to all the contralateral cortical neurons with equal probability, which suggests that each of these neurons needs to project to $\frac{n^2\rho_\mathrm{G}}{nf} = \frac{n\rho_\mathrm{G}}{f}$ neurons. A natural constraint is that $f \geqslant \rho_\mathrm{G}$.

An example of the $f = \rho_\mathrm{G}$ case is shown in Fig. S6(a): for $(m, n, \rho_\mathrm{W}, \rho_\mathrm{G}) = (20, 500, 0.1, 0.05)$, within one olfactory cortex, only $n \times f = 25$ neurons will project contralaterally and each of them has to project to all 500 neurons in the other olfactory cortex. However, even in this extreme case, the Hebbian solution $\mathbf{G}^\dagger$ matches well with the temporal mean $\langle \mathbf{G}_\mathrm{BA} \rangle$ [Fig. S6(b) and the pink dot in Fig. S6(c)], and the same for the analytical expression of test BAL Eq. (13) [the pink dot in Fig. S6(d)]. We also studied other values of $f$ from 0.1 to 1 with $\rho_\mathrm{G} = 0.05$ and found that the results do not depend on $f$ [see Figs. S6(c)–S6(d)], which is expected as long as all elements in $\hat{\mathbf{G}}_\mathrm{BA}, \mathbf{W}_\mathrm{A}, \mathbf{W}_\mathrm{B}$ are chosen independently (see SM Sec. II B [44] for details). These results indicate that significant alignment can be achieved even if only a small fraction of OC neurons project contralaterally.

### C. Global vs local learning rules: SGD modestly outperforms Hebbian rule but obeys the same inverse scaling relation

While the motivation of using the local Hebbian learning rule was to respect the biological constraint, we next asked whether global learning rules such as SGD, the predominant learning algorithm in artificial neural networks, could achieve better test BAL. To investigate this question, we introduced the following global alignment loss function:

$$\mathcal{L} = \frac{1}{2} \sum_{\mu=1}^{N} \left|\left| \mathbf{r}_\mathrm{B}^\mu - \lambda \mathbf{G}_\mathrm{BA} \mathbf{r}_\mathrm{A}^\mu \right|\right|^2, \tag{15}$$

where $\mu$ is the odor index and $\lambda$ is the scaling factor between ipsi- and contralateral responses. Other global objective functions such as the cosine similarity itself were also tested and did not affect the general results, as discussed in SM Sec. VII and shown by Fig. S7 [44]. The global loss function given in Eq. (15) is chosen for its simplicity and analytical tractability.

For each sequentially presented sample $\mathbf{x}^\mu$, the global "learning" dynamics is implemented by using the online SGD update rule:

$$\Delta \mathbf{G}_\mathrm{BA}^\mu = \eta \left( \mathbf{r}_\mathrm{B}^\mu \mathbf{r}_\mathrm{A}^{\mu\top} - \lambda \mathbf{G}_\mathrm{BA} \mathbf{r}_\mathrm{A}^\mu \mathbf{r}_\mathrm{A}^{\mu\top} \right) \odot \hat{\mathbf{G}}_\mathrm{BA}. \tag{16}$$

This SGD rule is nonlocal, because in the second term, $(\mathbf{G}_\mathrm{BA} \mathbf{r}_\mathrm{A}^\mu \mathbf{r}_\mathrm{A}^{\mu\top})_{ij} = \sum_k \mathbf{G}_{\mathrm{BA},ik} \mathbf{r}_{\mathrm{A},k}^\mu \mathbf{r}_{\mathrm{A},j}^\mu$ involves other synapses and neurons. Note that the first term is the same as in the Hebbian update rule Eq. (2), therefore, to make SGD and Hebbian updates (and their solutions) comparable in size, we adopted $\lambda = \frac{\beta}{m\rho_\mathrm{W}\gamma^2}$ (see SM Sec. VI [44] for details).

We found that with global information, SGD indeed outperforms Hebbian learning but has similar dependence on the network properties such as connection sparsity and network size. In particular, using the same parameters as the Hebbian learning simulations shown in Fig. 3(a), we varied $\rho_\mathrm{G}$ for different $n$ and computed the bilateral alignment level $a^\mathrm{S}$ for SGD. Quantitatively, we found that $a^\mathrm{S}(\rho_\mathrm{G}) > a^\mathrm{H}(\rho_\mathrm{G})$ [compare the solid (SGD) and the dashed (Hebbian) purple lines in

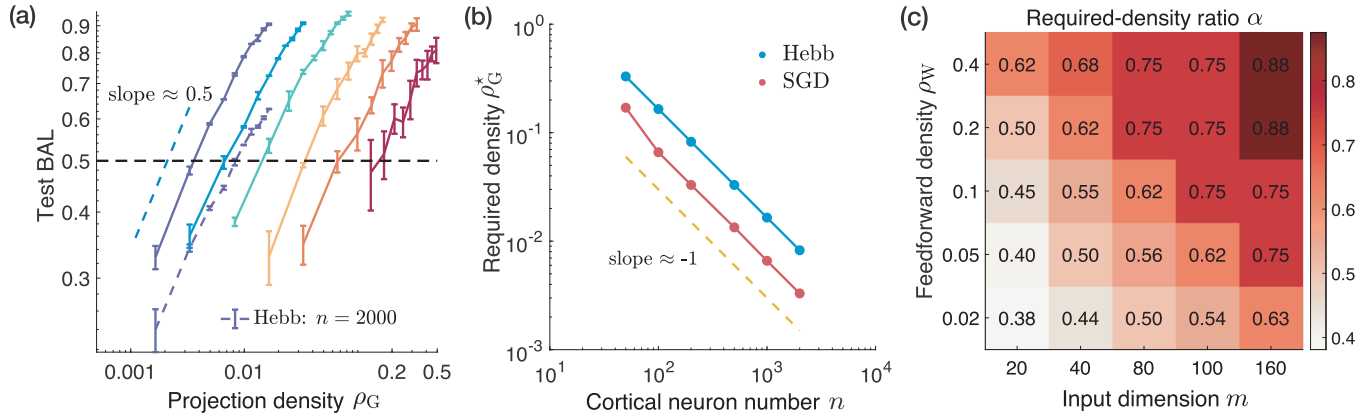


FIG. 4. Global SGD learning rule outperforms local Hebb's rule but has the same inverse scaling, and two learning rules get closer with larger $m$ and $\rho_W$. (a) Test BAL under SGD learning shows similar power-law scaling with $\rho_G$ as that in the Hebbian learning. Colors indicate different $n$ as in Fig. 3(c). (b) $\rho_G^\star$ in both Hebbian and SGD learning rules scale inversely with $n$. (c) Larger $m$ and $\rho_W$ lead to better alignment between Hebbian and SGD learning rules, as shown by the heatmap of $\alpha(m, \rho_W)$.

Fig. 4(a) for $n = 2000$]. However, $a^S$ has similar dependence on $\rho_G$ and $n$ as in Hebbian learning [Fig. 3(a)], e.g., the power-law relation between $a$ and $\rho_G$ for small $\rho_G$ with the same exponent 0.5 (the left blue dashed line). Furthermore, following the analysis in the previous section, we defined $\rho_G^{\star S}(n)$ as the density for SGD to achieve a given level of alignment $a^\star$, i.e., $a^S[\rho_G^{\star S}(n)] = a^\star$, and found the same inverse scaling between $\rho_G^{\star S}$ and $n$ [the red line in Fig. 4(b)].

Since Hebbian learning and SGD both exhibit the inverse scaling, the ratio of $\rho_G^{\star S}(n)$ to $\rho_G^{\star H}(n)$ is a constant. Therefore, to quantitatively compare Hebb's rule and SGD, we defined $\alpha(m, \rho_W) := \langle \frac{\rho_G^{\star S}(n)}{\rho_G^{\star H}(n)} \rangle_n$, the mean ratio of required SGD interhemispheric connection density to that of Hebbian learning. A smaller $\alpha$ means that SGD needs fewer interhemispheric connections to achieve the same performance as the Hebbian learning. For $m = 20$, $\rho_W = 0.1$ used in Fig. 3, we found that $\alpha \approx 0.42$ (Fig. S8), corresponding to modestly sparser connectivity. By varying $m$ and $\rho_W$, we found that $\alpha(m, \rho_W)$ increases with $m$ and $\rho_W$ [Fig. 4(c)] and approaches $\sim 1$ at large values of $m$ and $\rho_W$, meaning that Hebbian learning has almost the same performance as SGD, with more OB neurons and denser feedforward projections.

To explain the similarity between SGD and Hebb's rule for this alignment task, we compared their weight updates upon receiving the same input (training) data at the same time point along the learning trajectory. Specifically, the Hebbian and SGD updates at step $T$ are given by

$$\Delta \mathbf{G}_T^H = \eta\left(\mathbf{r}_B \mathbf{r}_A^\top - \beta \mathbf{G}_T^H\right) \odot \hat{\mathbf{G}}_{BA},$$
$$\Delta \mathbf{G}_T^S = \eta\left(\mathbf{r}_B \mathbf{r}_A^\top - \lambda \mathbf{G}_T^S \mathbf{r}_A \mathbf{r}_A^\top\right) \odot \hat{\mathbf{G}}_{BA}. \quad (17)$$

We calculated their cosine similarity, i.e., $\cos(\Delta \mathbf{g}_T^H, \Delta \mathbf{g}_T^S)$, where $\Delta \mathbf{g}$ is the vectorized $\Delta \mathbf{G}$. We found significant overlap between the two update vectors during the entire learning process as shown in Fig. 5(a). Furthermore, this update alignment level increases with $m$ and $\rho_W$ [see Fig. 5(b), where we averaged over $T$ and $n$ to get $\langle \cos(\Delta \mathbf{g}^H, \Delta \mathbf{g}^S) \rangle$]. We also calculated the solution alignment level, i.e., the cosine similarity between the final SGD solution at step $T = N$ and the Hebbian solution [$\langle \cos(\mathbf{g}^S, \mathbf{g}^\dagger) \rangle$ in Fig. 5(c)]. We found that the solution alignment level also increases with $m$ and $\rho_W$. In fact, both the solution alignment level and $\alpha$ are strongly correlated with the update alignment level [the blue and the red dots in Fig. 5(d)].

To understand the origin of the update alignment and its dependence on $m$ and $\rho_W$, we decomposed the weight update vectors $\mathbf{d}_1(\mathbf{x}) := \Delta \mathbf{g}_T^H$ and $\mathbf{d}_2(\mathbf{x}) := \Delta \mathbf{g}_T^S$ as

$$\mathbf{d}_1(\mathbf{x}) = \mathbf{h}(\mathbf{x}) - \mathbf{h}_1(\mathbf{x}),$$
$$\mathbf{d}_2(\mathbf{x}) = \mathbf{h}(\mathbf{x}) - \mathbf{h}_2(\mathbf{x}), \quad (18)$$

where $\mathbf{h}(\mathbf{x}) = \mathrm{vec}(\eta \mathbf{r}_B \mathbf{r}_A^\top \odot \hat{\mathbf{G}}_{BA})$ is the Hebbian correlation term shared by both update rules, $\mathbf{h}_1(\mathbf{x}) = \eta\beta \mathbf{g}_T^H$ is the Hebbian-specific update vector, and $\mathbf{h}_2(\mathbf{x}) = \mathrm{vec}(\eta\lambda \mathbf{G}_T^S \mathbf{r}_A \mathbf{r}_A^\top \odot \hat{\mathbf{G}}_{BA})$ is the SGD-specific update vector, as illustrated in Fig. 6(a). We then evaluated the averaged cosine similarity and amplitude ratio between the shared component $\mathbf{h}(\mathbf{x})$ and the specific components, $\mathbf{h}_1(\mathbf{x})$ and $\mathbf{h}_2(\mathbf{x})$, for Hebbian rule and SGD, respectively, which are denoted as

$$c_1 = \langle \cos(\mathbf{h}(\mathbf{x}), \mathbf{h}_1(\mathbf{x})) \rangle_{\mathbf{x}}, \quad r_1 = \langle |\mathbf{h}(\mathbf{x})| / |\mathbf{h}_1(\mathbf{x})| \rangle_{\mathbf{x}};$$
$$c_2 = \langle \cos(\mathbf{h}(\mathbf{x}), \mathbf{h}_2(\mathbf{x})) \rangle_{\mathbf{x}}, \quad r_2 = \langle |\mathbf{h}(\mathbf{x})| / |\mathbf{h}_2(\mathbf{x})| \rangle_{\mathbf{x}}. \quad (19)$$

We found that $r_1 \gg 1$, and it increases significantly with $m$ as shown in Fig. 6(b). This indicates that $\mathbf{d}_1(\mathbf{x})$ is highly aligned with $\mathbf{h}(\mathbf{x})$, and the alignment increases with $m$ as illustrated in Fig. 6(a) and shown in Fig. S9(b). For SGD, $r_2$ is larger than 1 and it increases with $m$ and $\rho_W$ as shown in Fig. 6(c), and $c_2 (< 1)$ decreases with $m$ and $\rho_W$ [see Fig. S9(c)]. Since $r_1 \gg 1$, we can approximate $\mathbf{d}_1 \approx \mathbf{h}$ and show that $\mathbf{d}_1 \cdot \mathbf{d}_2 \approx |\mathbf{h}|^2(1 - c_2/r_2) > 0$, which means that the Hebbian and SGD update vectors are positively aligned. Furthermore, as $r_2$ increases and $c_2$ decreases with $m$ and $\rho_W$, $\mathbf{d}_1(\mathbf{x})$ and $\mathbf{d}_2(\mathbf{x})$ align better with higher input dimension and denser feedforward projections as shown in Fig. 5(b). Consequently, a higher alignment between the update vectors of the two learning rules results in a higher similarity between the two solutions [Fig. 5(c)] and thus a larger required-density ratio $\alpha$ [Fig. 5(d)], which means that Hebbian learning is closer to SGD in bilateral alignment performance.

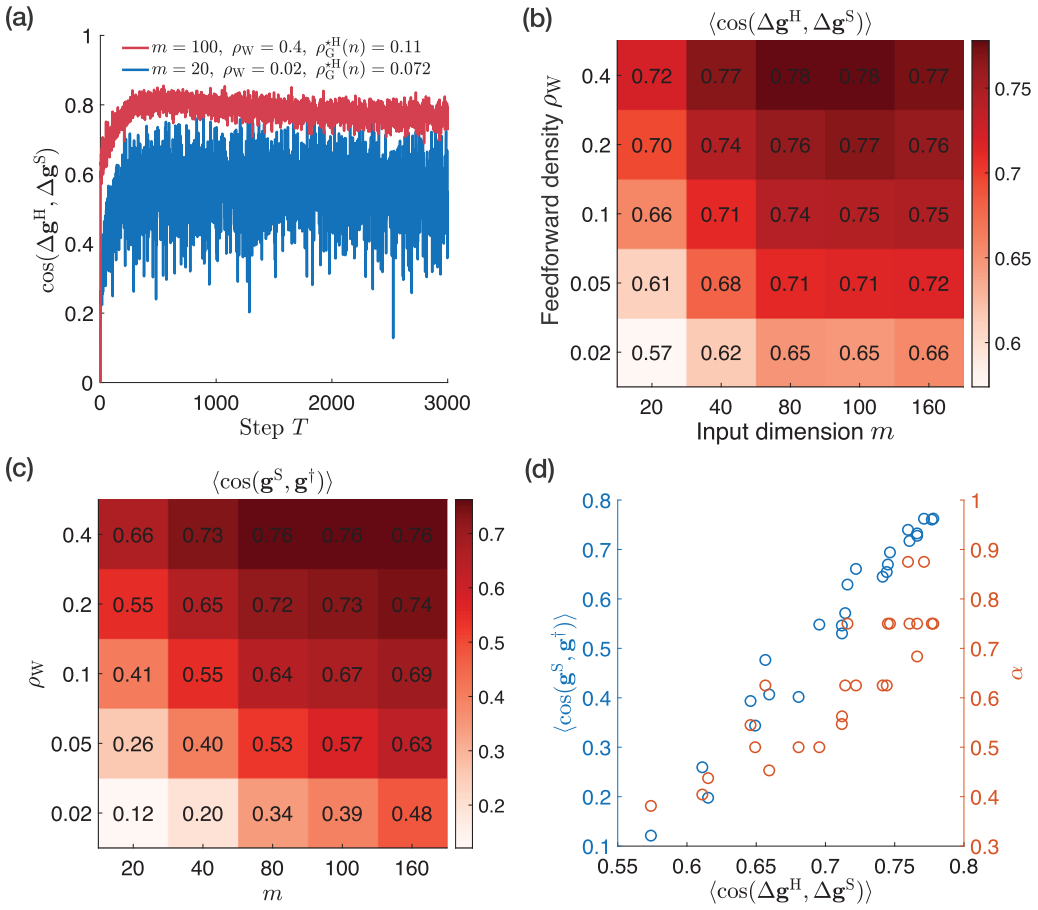


FIG. 5. The weight update vector of Hebbian learning aligns better with the corresponding SGD vector for larger $m$ and $\rho_W$, leading to higher solution alignment level and larger $\alpha$. (a) The temporal dynamics of update alignment $\cos(\Delta \mathbf{g}^H, \Delta \mathbf{g}^S)$ for two pairs of parameters [red for $m = 100$, $\rho_W = 0.4$, $\rho_G^{\star H}(n = 500) = 0.11$, and blue for $m = 40$, $\rho_W = 0.02$, $\rho_G^{\star H}(n = 500) = 0.072$]. (b) The update alignment level increases with $m$ and $\rho_W$. (c) Similar trend is observed for the final SGD solution and the theoretical Hebbian solution, i.e., $\cos(\mathbf{g}^S, \mathbf{g}^\dagger)$. (d) Both the solution alignment level $\cos(\mathbf{g}^S, \mathbf{g}^\dagger)$ (blue dots for the left $y$-axis) and $\alpha$ (red dots for the right $y$ axis) strongly correlate with update alignment level $\cos(\Delta \mathbf{g}^H, \Delta \mathbf{g}^S)$. Here each dot corresponds to one pair of ($m$, $\rho_W$) in panels (b) and (c).

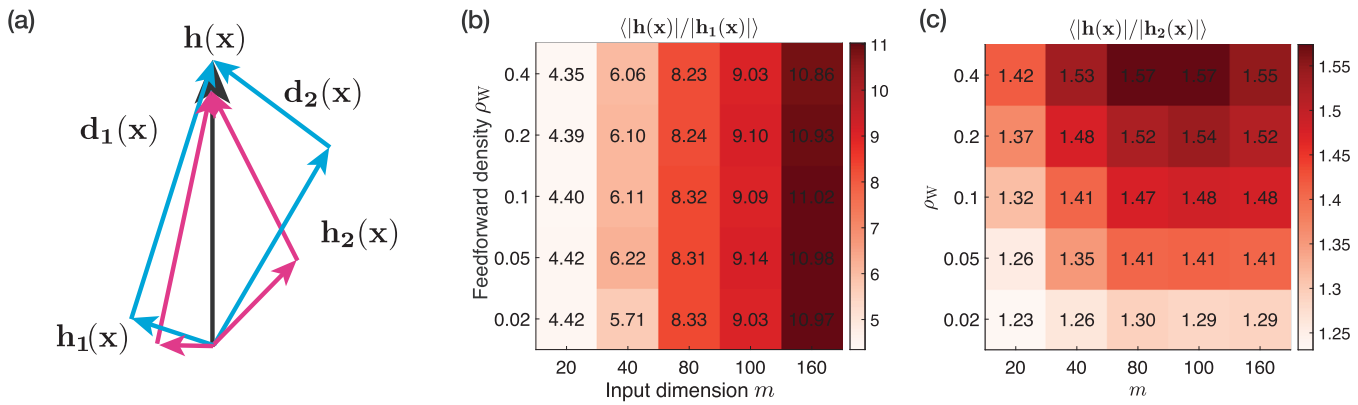


FIG. 6. Both Hebbian and SGD weight update vectors overlap more with the common Hebbian term with larger $m$ and $\rho_W$. (a) The geometric representation of how $\mathbf{h}_1(\mathbf{x})$, $\mathbf{h}_2(\mathbf{x})$, $\mathbf{d}_1(\mathbf{x})$, $\mathbf{d}_2(\mathbf{x})$ change with ($m$, $\rho_W$), compared to $\mathbf{h}(\mathbf{x})$ (the central arrow). Blue arrows: small $m$ and $\rho_W$. Pink arrows: large $m$ and $\rho_W$. (b) $|\mathbf{h}(\mathbf{x})|/|\mathbf{h}_1(\mathbf{x})| \gg 1$ and increases with $m$. Here, for each pair of ($m$, $\rho_W$), the mean is obtained by averaging over $\mathbf{x}$ and $n$. (c) In general, $\mathbf{h}(\mathbf{x})$ is more dominant over $\mathbf{h}_2(\mathbf{x})$ in magnitude for larger $m$ and $\rho_W$.

## IV. DISCUSSION

We have demonstrated that a simple online Hebbian learning rule is sufficient to align bilateral cortical representations even with sparse interhemispheric projections. Through numerical simulations and analytical calculations of a simplified neural network model, we discovered that to achieve a certain level of bilateral alignment, each cortical neuron (on average) must receive a fixed number of interhemispheric projections that is independent of the total number of cortical neurons. Consequently, the required projection density scales inversely with the number of cortical neurons. Remarkably, this scaling law also holds for a global learning rule, SGD, that optimizes a similar alignment loss function. Although SGD outperforms the local Hebbian rule in terms of achieving higher bilateral alignment levels, the difference diminishes with higher input dimensions and denser feedforward projections. Our analysis revealed that this reduction is due to a higher overlap between the Hebbian update vector and the SGD update vector.

Our model offers a mechanistic understanding of how the olfactory system achieves bilateral alignment, a mechanism that could extend to other neural systems. Additionally, our model generates several predictions or insights that are potentially testable. Firstly, the inverse scaling relation between the cortical neuron number and the interhemispheric connectivity could be be examined through analysis of connectomic data across different species and brain regions. In addition to this inverse scaling, Eq. (14) predicts other quantitative and testable relations, specifically, given $n$ and $a^\star$, the required density $\rho_{\rm G}^\star$ increases with $m$ and $1/\rho_{\rm W}$. Finally, our model predicts that the fine structure of OC (the fraction of subpopulation projecting contralaterally) does not affect the alignment performance or the inverse scaling.

Our work may also have implications for other brain areas where representations in the two hemispheres have to be matched or integrated. In sensory regions with topographic representations, a coarse spatial template for organizing cross-hemispheric projections can potentially be genetically specified, with further refinements arising through activity-dependent mechanisms [49–51]. However, in other brain regions with more scattered, mosaic representations, coordination is likely to occur through more flexible activity-dependent mechanisms. Commissural projections in the medial entorhinal cortex, an area strongly implicated in spatial and episodic memory, play an active role in memory retrieval [52]. Whether this interhemispheric interaction involves detailed functional alignment is an interesting question for future work. In another intriguing example, interhemispheric projections in a brain region called the anterior lateral motor cortex has been shown to play a role in working memory by helping restore population activity when perturbed [53,54]. Remarkably, this restoration is a key hallmark of attractor networks, which may be formed flexibly during learning, potentially through Hebbian-like mechanisms in interhemispheric connections [54,55]. Therefore, our modeling and analytical framework could be extended to these other regions and functions.

The update alignment between SGD and Hebb's rule shares the same feature as many recently developed, biologically plausible learning algorithms for training deep neural networks, such as the feedback alignment algorithm and its variants [56,57]. There, the learning signals generally aligns well with that of the backpropagation algorithm (BP) [58,59]. In our bilateral alignment task, the local and unsupervised Hebb's rule exhibits overlap with the weight update vector of SGD, which optimizes a global alignment loss function. These results suggest that a biologically plausible learning algorithm can perform well if its weight update contains information regarding the corresponding error gradient. This insight provides a guidance to search for biologically plausible learning algorithms in a much larger admissible solution space. This is because the algorithm does not have to be an approximation of the BP as long as its learning signal has significant overlap with that in the BP. Future studies in comparing learning dynamics with the Hebbian-like rule and the global SGD rule are necessary for developing brain-inspired efficient local learning algorithms to solve more complex problems.

Besides the alignment of the weight update vectors between Hebbian and SGD rules, sparse connectivity as reflected in the shared inverse scaling law is another common theme in biological and artificial neural networks. Sparse connectivity has been widely observed in the brain [60,61]. For artificial neural networks, sparse connectivity is critical for enhancing algorithm efficiency, and it is often achieved by pruning trained networks [62–68] with some of the pioneering works there partially inspired by neuroscience [63,64]. In recent years, various methods and applications have been developed to prune and sparsify large language models [69–73]. Despite their differences, in both biological and artificial neural networks, the common goal of obtaining sparse networks is to reduce the required computational resources. How to minimize the number of parameters by sparsifying the network without affecting its performance remains a major challenge for both biological and artificial neural networks.

We end the paper by pointing out that our current model, aimed at simplicity and analytical tractability, inevitably simplifies biological complexity and thus entails certain limitations. We discuss some of the limitations and possible future directions to address them in the following.

Firstly, for the training data used in our model for alignment, we opted for IID Gaussian input to facilitate analytical tractability. However, realistic OB signals may be correlated [74]. Our preliminary results with simple correlated inputs showed that the critical $\rho_G^*$ is lowered for correlated inputs, presumably due to the effectively lower input dimension. However, the inverse scaling does not change (SM Sec. IV and Fig. S10 [44]). It remains an interesting direction to study whether and how statistics of more complex inputs affect the alignment performance [Eq. (13)] and the inverse scaling.

Secondly, our model for the OB-OC projections is simplified as it included only the "effective" direct connections $\mathbf{W}_{\rm A}$ and $\mathbf{W}_{\rm B}$. For analytical tractability, elements of $\mathbf{W}_{\rm A}$ and $\mathbf{W}_{\rm B}$ take randomly assigned positive or negative values, which seemingly violates Dale's principle [75]. However, such simplification may be justified by considering the indirect inhibitory connections through the feed-forward inhibitory neurons in the OC [48,76,77], which can lead to odor-evoked suppressed cortical responses [36,78]. In addition, our work adopted random OB-to-OC projections, but a recent work showed that these projections exhibit some structure [79].

Studying how local inhibitory circuits and other structural constraints affect distributions of $\mathbf{W}_A$ and $\mathbf{W}_B$ and the bilateral alignment performance provides another avenue for future exploration.

Last but not least, our current model did not explicitly incorporate the recurrent connections within each olfactory cortex [48,76]. Such recurrent connections have been postulated to play a significant role in associative memory and pattern completion: with distorted inputs it can recover the true patterns through internal dynamics [80–83]. But how such recurrent connections affect the bilateral alignment remains an interesting open question. As an initial step, we developed a minimum model for alignment with recurrent connections in SM Sec. V [44] and tested two specific choices of recurrent connections (Hebbian-learned and random). We found that although including sparse recurrent connections with strength similar to that of interhemispheric connections did change the response of individual neurons as expected, they do not change the alignment performance significantly, and the inverse scaling remains valid [Figs. S11–S12]. More systematic studies that take into account realistic structure and dynamics of the recurrent connections need to be done in order to understand how recurrent connections affect alignment performance.

## ACKNOWLEDGMENTS

We thank Cengiz Pehlevan for helpful discussions, and Julien Grimaud for the inspiring experiments and for sharing the illustrations. The work of Y.T. was partially supported by an NIH grant (Grant No. R35GM131734). This work was supported by a Harvard Mind Brain Behavior Graduate Student Award and a grant from NTT Research, and was partially done when B.L. was a summer intern at IBM Research. Experiments related to this work were supported in part by an NIH grant (Grant No. RF1NS128865) to V.N.M., and Y.T. would like to thank the Center for Computational Biology at the Flatiron Institute for hospitality while a portion of this work was carried out.

[1] B. Grothe and M. Pecka, The natural history of sound localization in mammals-A story of neuronal inhibition, Front. Neural Circuits **8**, 116 (2014).

[2] D. McAlpine, Binaural pathways and processing, in *Encyclopedia of Neuroscience*, edited by M. D. Binder, N. Hirokawa, and U. Windhorst (Springer, Berlin, Heidelberg, 2009), pp. 383–388.

[3] R. Blake and R. Fox, The psychophysical inquiry into binocular summation, Percept. Psychophys. **14**, 161 (1973).

[4] B. G. Cumming and G. C. DeAngelis, The physiology of stereopsis, Annu. Rev. Neurosci. **24**, 203 (2001).

[5] R. Rajan, J. P. Clement, and U. S. Bhalla, Rats smell in stereo, Science **311**, 666 (2006).

[6] J. Porter, T. Anand, B. Johnson, R. M. Khan, and N. Sobel, Brain mechanisms for extracting spatial information from smell, Neuron **47**, 581 (2005).

[7] J. D. Mainland, E. A. Bremner, N. Young, B. N. Johnson, R. M. Khan, M. Bensafi, and N. Sobel, One nostril knows what the other learns, Nature (London) **419**, 802 (2002).

[8] J. E. Rabell, K. Mutlu, J. Noutel, P. M. Del Olmo, and S. Haesler, Spontaneous rapid odor source localization behavior requires interhemispheric communication, Curr. Biol. **27**, 1542 (2017).

[9] T. Dalal, N. Gupta, and R. Haddad, Bilateral and unilateral odor processing and odor perception, Commun. Biol. **3**, 150 (2020).

[10] S. Kikuta, H. Kashiwadani, and K. Mori, Compensatory rapid switching of binasal inputs in the olfactory cortex, J. Neurosci. **28**, 11989 (2008).

[11] D. A. Wilson and R. M. Sullivan, Respiratory airflow pattern at the rat's snout and an hypothesis regarding its role in olfaction, Physiol. Behav. **66**, 41 (1999).

[12] R. Eccles, Nasal airflow in health and disease, Acta Oto-Laryngol. **120**, 580 (2000).

[13] K. C. Catania, Stereo and serial sniffing guide navigation to an odour source in a mammal, Nat. Commun. **4**, 1441 (2013).

[14] D. H. Gire, V. Kapoor, A. Arrighi-Allisan, A. Seminara, and V. N. Murthy, Mice develop efficient strategies for foraging and navigation using complex natural stimuli, Curr. Biol. **26**, 1261 (2016).

[15] U. Bhattacharyya and U. S. Bhalla, Robust and rapid air-borne odor tracking without casting, Eneuro **2**, ENEURO.0102-15.2015 (2015).

[16] T. M. Findley, D. G. Wyrick, J. L. Cramer, M. A. Brown, B. Holcomb, R. Attey, D. Yeh, E. Monasevitch, N. Nouboussi, I. Cullen *et al.*, Sniff-synchronized, gradient-guided olfactory search by freely moving mice, Elife **10**, e58523 (2021).

[17] Y. Wu, K. Chen, Y. Ye, T. Zhang, and W. Zhou, Humans navigate with stereo olfaction, Proc. Natl. Acad. Sci. **117**, 16065 (2020).

[18] D. Kucharski and W. Hall, New routes to early memories, Science **238**, 786 (1987).

[19] D. Kucharski, N. Burka, and W. Hall, The anterior limb of the anterior commissure is an access route to contralateral stored olfactory preference memories, Psychobiology **18**, 195 (1990).

[20] P. C. Brunjes, K. R. Illig, and E. A. Meyer, A field guide to the anterior olfactory nucleus (cortex), Brain Res. Rev. **50**, 305 (2005).

[21] S. Kikuta, K. Sato, H. Kashiwadani, K. Tsunoda, T. Yamasoba, and K. Mori, Neurons in the anterior olfactory nucleus pars externa detect right or left localization of odor sources, Proc. Natl. Acad. Sci. USA **107**, 12363 (2010).

[22] D. A. Wilson, Binaral interactions in the rat piriform cortex, J. Neurophysiol. **78**, 160 (1997).

[23] Z. Yan, J. Tan, C. Qin, Y. Lu, C. Ding, and M. Luo, Precise circuitry links bilaterally symmetric olfactory maps, Neuron **58**, 613 (2008).

[24] A. Hagiwara, S. K. Pal, T. F. Sato, M. Wienisch, and V. N. Murthy, Optophysiological analysis of associational circuits in the olfactory cortex, Front. Neural Circuits **6**, 18 (2012).

[25] L. B. Haberly and J. L. Price, Association and commissural fiber systems of the olfactory cortex of the rat, I. Systems originating in the piriform cortex and adjacent areas, J. Comp. Neurol. **178**, 711 (1978).

[26] M. J. Russo, K. M. Franks, R. Oghaz, R. Axel, and S. A. Siegelbaum, Synaptic organization of anterior olfactory nucleus inputs to piriform cortex, J. Neurosci. **40**, 9414 (2020).

[27] E. Martin-Lopez, S. J. Meller, and C. A. Greer, Development of piriform cortex interhemispheric connections via the anterior commissure: Progressive and regressive strategies, Brain Struct. Funct. **223**, 4067 (2018).

[28] G. N. Dikeçligil, A. I. Yang, N. Sanghani, T. Lucas, H. I. Chen, K. A. Davis, and J. A. Gottfried, Odor representations from the two nostrils are temporally segregated in human piriform cortex, Curr. Biol. **33**, 5275 (2023).

[29] L. B. Haberly, Parallel-distributed processing in olfactory cortex: New insights from morphological and physiological analysis of neuronal circuitry, Chem. Senses **26**, 551 (2001).

[30] D. L. Sosulski, M. L. Bloom, T. Cutforth, R. Axel, and S. R. Datta, Distinct representations of olfactory information in different cortical centres, Nature (London) **472**, 213 (2011).

[31] K. Miyamichi, F. Amat, F. Moussavi, C. Wang, I. Wickersham, N. R. Wall, H. Taniguchi, B. Tasic, Z. J. Huang, Z. He *et al.*, Cortical representations of olfactory input by trans-synaptic tracing, Nature (London) **472**, 191 (2011).

[32] S. Ghosh, S. D. Larson, H. Hefzi, Z. Marnoy, T. Cutforth, K. Dokka, and K. K. Baldwin, Sensory maps in the olfactory cortex defined by long-range viral tracing of single neurons, Nature (London) **472**, 217 (2011).

[33] E. S. Schaffer, D. D. Stettler, D. Kato, G. B. Choi, R. Axel, and L. Abbott, Odor perception on the two sides of the brain: Consistency despite randomness, Neuron **98**, 736 (2018).

[34] A. Celani, E. Villermaux, and M. Vergassola, Odor landscapes in turbulent environments, Phys. Rev. X **4**, 041015 (2014).

[35] G. Reddy, V. N. Murthy, and M. Vergassola, Olfactory sensing and navigation in turbulent environments, Annu. Rev. Condens. Matter Phys. **13**, 191 (2022).

[36] J. Grimaud, W. Dorrell, S. Jayakumar, C. Pehlevan, and V. Murthy, Bilateral alignment of receptive fields in the olfactory cortex, eNeuro **11**, ENEURO.0155-24.2024 (2024).

[37] S. Srinivasan and C. F. Stevens, Scaling principles of distributed circuits, Curr. Biol. **29**, 2533 (2019).

[38] B. G. Cragg, The density of synapses and neurones in the motor and visual areas of the cerebral cortex, J. Anat. **101**, 639 (1967).

[39] A. Schüz and G. Palm, Density of neurons and synapses in the cerebral cortex of the mouse, J. Comp. Neurol. **286**, 442 (1989).

[40] L. Wang, Z. Zhang, J. Chen, A. Manyande, R. Haddad, Q. Liu, and F. Xu, Cell-type-specific whole-brain direct inputs to the anterior and posterior piriform cortex, Front. Neural Circuits **14**, 4 (2020).

[41] V. N. Murthy, Olfactory maps in the brain, Annu. Rev. Neurosci. **34**, 233 (2011).

[42] N. Uchida, C. Poo, and R. Haddad, Coding and transformations in the olfactory system, Annu. Rev. Neurosci. **37**, 363 (2014).

[43] B. Babadi and H. Sompolinsky, Sparseness and expansion in sensory representations, Neuron **83**, 1213 (2014).

[44] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PRXLife.2.043016 for the derivations and methods (Secs. I - IX), parameters (Table I), and Figs. S1 - S12.

[45] G. E. Uhlenbeck and L. S. Ornstein, On the theory of the Brownian motion, Phys. Rev. **36**, 823 (1930).

[46] S. Srinivasan and C. F. Stevens, The distributed circuit within the piriform cortex makes odor discrimination robust, J. Comp. Neurol. **526**, 2725 (2018).

[47] M. B. Richard, S. R. Taylor, and C. A. Greer, Age-induced disruption of selective olfactory bulb synaptic circuits, Proc. Natl. Acad. Sci. **107**, 15613 (2010).

[48] J. M. Bekkers and N. Suzuki, Neurons and circuits for odor processing in the piriform cortex, Trends Neurosci. **36**, 429 (2013).

[49] H. Mizuno, T. Hirano, and Y. Tagawa, Evidence for activity-dependent cortical wiring: Formation of interhemispheric connections in neonatal mouse visual cortex requires projection neuron activity, J. Neurosci. **27**, 6760 (2007).

[50] R. Suarez, L. R. Fenlon, R. Marek, L. Avitan, P. Sah, G. J. Goodhill, and L. J. Richards, Balanced interhemispheric cortical activity is required for correct targeting of the corpus callosum, Neuron **82**, 1289 (2014).

[51] K.-S. Lee, K. Vandemark, D. Mezey, N. Shultz, and D. Fitzpatrick, Functional synaptic architecture of callosal inputs in mouse primary visual cortex, Neuron **101**, 421 (2019).

[52] A. Caputi, X. Liu, E. C. Fuchs, Y.-C. Liu, and H. Monyer, Medial entorhinal cortex commissural input regulates the activity of spatially and object-tuned cells contributing to episodic memory, Neuron **110**, 3389 (2022).

[53] N. Li, K. Daie, K. Svoboda, and S. Druckmann, Robust neuronal dynamics in premotor cortex during motor planning, Nature (London) **532**, 459 (2016).

[54] G. Chen, B. Kang, J. Lindsey, S. Druckmann, and N. Li, Modularity and robustness of frontal cortical networks, Cell **184**, 3717 (2021).

[55] H. K. Inagaki, L. Fontolan, S. Romani, and K. Svoboda, Discrete attractor dynamics underlies persistent activity in the frontal cortex, Nature (London) **566**, 212 (2019).

[56] T. P. Lillicrap, D. Cownden, D. B. Tweed, and C. J. Akerman, Random synaptic feedback weights support error backpropagation for deep learning, Nat. Commun. **7**, 13276 (2016).

[57] A. Nøkland, Direct feedback alignment provides learning in deep neural networks, *NIPS'16: Proceedings of the 30th International Conference on Neural Information Processing Systems* (ACM, New York, 2016), pp. 1045–1053.

[58] M. Refinetti, S. D'Ascoli, R. Ohana, and S. Goldt, Align, then memorise: The dynamics of learning with feedback alignment, in *International Conference on Machine Learning* (PMLR, 2021), Vol. 139, pp. 8925–8935.

[59] J. Launay, I. Poli, and F. Krzakala, Principled training of neural networks with direct feedback alignment, arXiv:1906.04554.

[60] D. Meunier, R. Lambiotte, and E. T. Bullmore, Modular and hierarchically modular organization of brain networks, Front. Neurosci. **4**, 200 (2010).

[61] L. Pessoa, Understanding brain networks and brain organization, Phys. Life Rev. **11**, 400 (2014).

[62] Sietsma and Dow, Neural net pruning—Why and how, in *IEEE 1988 International Conference on Neural Networks* (IEEE, Piscataway, NJ, 1988), pp. 325–333.

[63] Y. LeCun, J. Denker, and S. Solla, Optimal brain damage, *Advances in Neural Information Processing Systems 2 (NIPS 1989)* (ACM, New York, NY, 1989).

[64] B. Hassibi and D. G. Stork, Second order derivatives for network pruning: Optimal brain surgeon,

*Advances in Neural Information Processing Systems* (ACM, New York, NY, 1992), Vol. 5.

[65] R. Reed, Pruning algorithms-a survey, IEEE Trans. Neural Networks **4**, 740 (1993).

[66] S. Han, J. Pool, J. Tran, and W. Dally, Learning both weights and connections for efficient neural network, *Advances in Neural Information Processing Systems* (ACM, New York, NY, 2015), Vol. 28.

[67] J. Frankle and M. Carbin, The lottery ticket hypothesis: Finding sparse, trainable neural networks, arXiv:1803.03635.

[68] T. Hoefler, D. Alistarh, T. Ben-Nun, N. Dryden, and A. Peste, Sparsity in deep learning: Pruning and growth for efficient inference and training in neural networks, J. Mach. Learn. Res. **22**, 1 (2021).

[69] X. Ma, G. Fang, and X. Wang, LLM-Pruner: On the structural pruning of large language models, *Advances in Neural Information Processing Systems* (ACM, New York, NY, 2023).

[70] M. Sun, Z. Liu, A. Bair, and J. Z. Kolter, A simple and effective pruning approach for large language models, arXiv:2306.11695.

[71] M. Xia, T. Gao, Z. Zeng, and D. Chen, Sheared LLaMA: Accelerating language model pre-training via structured pruning, arXiv:2310.06694.

[72] B. Liu and Y. Xu, Sparsity limit to prune large language models for on-device AI assistants: Llama-2 as an example, Preprints 202407.1568.v2 (2024).

[73] B. Liu, Comparing sparse Llama 3 and Llama 2 models for on-device AI assistants, Research Square rs.3.rs-4927672/v2 (2024).

[74] L. Ma, Q. Qiu, S. Gradwohl, A. Scott, E. Q. Yu, R. Alexander, W. Wiegraebe, and C. R. Yu, Distributed representation of chemical features and tunotopic organization of glomeruli in the mouse olfactory bulb, Proc. Natl. Acad. Sci. **109**, 5481 (2012).

[75] J. C. Eccles, From electrical to chemical transmission in the central nervous system: The closing address of the Sir Henry Dale Centennial Symposium Cambridge, 19 September 1975, Notes Rec. R. Soc. Lond. **30**, 219 (1976).

[76] K. M. Franks, M. J. Russo, D. L. Sosulski, A. A. Mulligan, S. A. Siegelbaum, and R. Axel, Recurrent circuitry dynamically shapes the activation of piriform cortex, Neuron **72**, 49 (2011).

[77] M. Stern, K. A. Bolding, L. Abbott, and K. M. Franks, A transformation from temporal to ensemble coding in a model of piriform cortex, eLife **7**, e34831 (2018).

[78] C. Poo and J. S. Isaacson, Odor representations in olfactory cortex: "Sparse" coding, global inhibition, and oscillations, Neuron **62**, 850 (2009).

[79] Y. Chen, X. Chen, B. Baserdem, H. Zhan, Y. Li, M. B. Davis, J. M. Kebschull, A. M. Zador, A. A. Koulakov, and D. F. Albeanu, High-throughput sequencing of single neuron projections reveals spatial organization in the olfactory cortex, Cell **185**, 4117 (2022).

[80] J. J. Hopfield, Neural networks and physical systems with emergent collective computational abilities, Proc. Natl. Acad. Sci. **79**, 2554 (1982).

[81] L. B. Haberly and J. M. Bower, Olfactory cortex: Model circuit for study of associative memory? Trends Neurosci. **12**, 258 (1989).

[82] K. A. Bolding, S. Nagappan, B.-X. Han, F. Wang, and K. M. Franks, Recurrent circuitry is required to stabilize piriform cortex odor representations across brain states, Elife **9**, e53125 (2020).

[83] S. L. Pashkovski, G. Iurilli, D. Brann, D. Chicharro, K. Drummey, K. M. Franks, S. Panzeri, and S. R. Datta, Structure and flexibility in cortical representations of odour space, Nature (London) **583**, 253 (2020).