bioRxiv preprint doi: https://doi.org/10.1101/2023.02.28.530474; this version posted March 1, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

In vivo ephaptic coupling allows memory network formation

Dimitris A. Pinotsis[1,2] and Earl K. Miller[2]

[1]Centre for Mathematical Neuroscience and Psychology and Department of Psychology, City —University of London, London EC1V 0HB, United Kingdom

[2]The Picower Institute for Learning & Memory and Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

Correspondence: Dimitris A. Pinotsis

Centre for Mathematical Neuroscience and Psychology and Department of Psychology, City —University of London, London EC1V 0HB, United Kingdom
pinotsis@mit.edu

KEYWORDS: memory engrams; neural ensembles; working memory; synergetics; predictive coding; auto-encoders; effective connectivity

Acknowledgements. This work is supported by UKRI ES/T01279X/1, Office of Naval Research N00014-22-1-2453, The JPB Foundation and The Picower Institute for Learning and Memory.

bioRxiv preprint doi: https://doi.org/10.1101/2023.02.28.530474; this version posted March 1, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Abstract

It is increasingly clear that memories are distributed across multiple brain areas. Such "engram complexes" are important features of memory formation and consolidation. Here, we test the hypothesis that engram complexes are formed in part by bioelectric fields that sculpt and guide the neural activity and tie together the areas that participate in engram complexes. Like the conductor of an orchestra, the fields influence each musician or neuron and orchestrate the output, the symphony. Our results use the theory of synergetics, machine learning and data from a spatial delayed saccade task and provide evidence for in vivo ephaptic coupling in memory representations.

Introduction

In recent decades there has been a paradigm shift in neuroscience. In the past, we focused on properties of individual neurons[1,2]. There is now a growing realization that information storage and processing depends on spatially distributed, dynamic groupings of neurons [3,5,6], known as neural ensembles [7–11] or engram cells[12,13] . Techniques like protein induction[14], immediate early gene (IEG) expression [15] and optogenetics [16] allow for identification of ensemble neurons participating in memory storage and recall [17,18]. Further, recent experiments have found simultaneous neural ensembles maintaining the same memory in many brain areas, something known as engram complex [19,20]. In [19] a total of 247 brain areas were mapped using the protein cFos and IEG. Among them, 117 areas were found to be significantly reactivated when a fear memory was recalled. Thus, memory was not stored in a single brain area but was dispersed in multiple areas and neural ensembles. Earlier theories like memory consolidation[21] and multiple traces [22]have also found that memories are stored in multiple areas forming engram complexes. These are connected via engram pathways formed by mono- or poly-synaptic connections [23].

The challenge, then, is in understanding how the brain forms engram complexes. Each brain area is connected to many others. Anatomical connectivity alone cannot be the whole story. Hypotheses that could explain this include that engram complexes are dynamically formed by emergent properties of neurons like synchronized rhythms[24–27], possibly resulting from internal coordination of spike timing[28,29], that allow neuronal communication,[30–32] feature integration and perceptual segmentation[33,34]. Here, we report tests of the hypothesis that the

bioRxiv preprint doi: https://doi.org/10.1101/2023.02.28.530474; this version posted March 1, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

electric fields generated by neurons play a crucial role. We suggest that ephaptic coupling[35,36] ties together the areas that participate in engram complexes. In other words, we test the hypothesis that memory networks include electric fields that carry information back to individual neurons.

Direct evidence of ephaptic coupling of spiking has been found in brain slices[37–39]. *In vitro* ephaptic coupling has been found in LFPs. Application of external electric fields resulted in membrane potentials oscillating at the same frequency as the drive[35]. Support for its role in forming engram complexes comes from studies showing that neurons participating in an engram complex showed similar functional connectivity during optogenetic activation and memory recall[19],[40]. We found that the electric fields in the primate prefrontal cortex carried information about the contents of working memory[41]. Using data from a delayed saccade task[10,42], we built two models: one for neural activity [9,10] and another for the emergent electric field. This revealed electric field patterns that varied with contents of working memory. Further, we found that the electric fields were robust and stable while neural activity underlying memory showed representational drift. This latter observation suggested the hypothesis that electric fields could act as "guard rails" that help stabilize and funnel the high dimensional variable neural activity along stable lower-dimensional routes.

Here we test the hypothesis that electric fields sculpt and guide the neural activity forming engram complexes. We used a theory of complex systems known as synergetics[43,44]. We also extended the single area analysis of [41] and focused on data from two areas known to form an engram complex, Frontal Eye Fields (FEF) and Supplementary Eye Fields (SEF). Synergetics describes how complex systems (e.g. molecules, fluids, brain etc.) self – organize. In the case of human behavior, synergetics describe how the collective dynamics of muscles and body parts (e.g. fingers) give rise to behavior like rhythmic hand movement[45]. We applied synergetics to understand the emergence of memory representations. We performed mathematical, i.e. pen and paper, computations and showed that the theory predicts that electric fields guide ensemble activity. If ephaptic coupling occurs in a brain area and this exchanges memory information with other brain areas, then ephaptic coupling will occur in those areas too. We then confirmed our results using Bayesian Model Comparison[46,47], Granger Causality[48] and Representation Similarity Analysis[49]. This suggested that the electric field enslaves neurons, not the other way around. Applying the slaving principle[43], we found that the electric field controls neural activity and oscillations

bioRxiv preprint doi: https://doi.org/10.1101/2023.02.28.530474; this version posted March 1, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

through ephaptic coupling[39,50] and that this was the case across all recording sites that participated in the engram complex.

Methods

Task and Experimental Setup. We reanalyzed data from [42]. The same data were used in our earlier papers [51,52]. Two adult male Macaca monkeys were trained to perform an oculomotor spatial delayed response task. This task required the monkeys to maintain the memory of a randomly chosen visual target (at angles of 0, 60, 120, 180, 240 and 300 degrees, 12.5-degree eccentricity) over a brief (750 ms) delay period and then saccade to the remembered location. If a saccade was made to the cued angle, the target was presented with a green highlight and a water reward was delivered. If not, the target was presented with a red highlight and reward was withheld. 32-electrode chronic arrays were implanted unilaterally in FEF and SEF in each monkey. Each array consisted of a 2 x 2 mm square grid, where the spacing between electrodes was 400 um. The implant channels were determined prior to surgery using structural magnetic resonance imaging and anatomical atlases. From each electrode, we acquired local field potentials (LFPs; extracted with a fourth order Butterworth low-pass filter with a cut-off frequency of 500Hz, and recorded at 1 kHz) using a multichannel data acquisition system (Cerebus, Blackrock Microsystems). We analyzed LFPs during the delay period when monkeys held the cued angles in memory.

Deep neural fields describe neural ensemble activity

In previous work [41,51], we used a neural field model [53–56] to describe the evolution of the ensemble transmembrane potential or depolarization, V^m, in neural ensembles. Neural ensembles are groups of neurons that maintain working memory representations. Remembered stimuli can include locations, colours, categories or rules. In grey matter, current in a neural ensemble flows along the neurons' axons and dendrites. Chemical energy is converted to electrical. Action and synaptic potentials are summed up to produce an emerging *electric potential* (EP) V^e in extracellular space. The difference of intracellular

bioRxiv preprint doi: https://doi.org/10.1101/2023.02.28.530474; this version posted March 1, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

V^i and extracellular V^e potentials, $V^m = V_0^e - V_0^i$ is the transmembrane potential. This is described by neural network equations, like the Wilson—Cowan equations[41,57,58]

$$\dot{V}_X^m(x,y,t) = -\tau_X V_X^m(x,y,t) + \sum K_{XX}(x,y,x',y') f[V_X^m(x',y',t)] + \\ + \sum K_{XX'}(x,y,x'',y'') f[V_{X'}^m(x'',y'',t)] + U, \quad (1)$$

Here, x and y parameterize the location on a cortical patch occupied by the ensemble, X is an index denoting excitatory or inhibitory populations, K is the connectivity or weight matrix that describes how the signal is amplified or attenuated when it propagates between recording sites and $f(h) = \frac{1}{1+\exp(\delta(\eta - h))}$ is called transfer function and predicts average firing rate, similar to activation functions in deep neural networks [59,60]. Also, τ_X is the rate-constant of postsynaptic filtering, δ is synaptic gain and η is the postsynaptic potential at which the half of the maximum firing rate is achieved, see e.g. [61] for more details. In [41], starting from Equations (1), we obtained a neural field model

$$\dot{V}^m = -\tau_{NA} V^m + K * f \circ V^m + U \\ K * Q = \iint K(\upsilon, \psi, t, t') \cdot Q(\psi, t') d\psi dt' \quad (2)$$

where ψ is the location where afferent input originates. In [41,51], we rewrote the neural field model given by Equations (2) as a Gaussian Linear Model

bioRxiv preprint doi: https://doi.org/10.1101/2023.02.28.530474; this version posted March 1, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

$$
\begin{aligned}
Y &= \sum_j H_j w_j + m + \varepsilon \\
w &= \left[A_0, A_1, A_2, \ldots, A_{i-1}, A_i \right]^T \\
m &= N^{-1} \sum_N V^{ml}
\end{aligned}
\tag{3}
$$

where $\varepsilon \sim (0, s_s{}^2 I)$ and $H = \left[G_0, G_1, G_2, \ldots \right]^T$ are called principal axes and are given by the spatial derivatives of the neural activity. We will return to the principal axes below. Also, w is the vector of connectivity components. For more details, see [41,51]. The term "*deep*" was used to distinguish this model (with learned connectivity parameters) from common neural field models where connectivity weights are chosen *ad hoc*. These can be found by training them deep neural field model using an autoencoder algorithm, that is, by optimizing a cost function known as the Free Energy, F,

$$
\begin{aligned}
F &= \left(-\frac{1}{2} \right) \left[(Y - Hw)^T r_s^2 (Y - Hw) + \ln \left| s_s^2 \right| + \ln \left| s^2{}_s \Delta^{-1} \right| + Z^T Z + \mathrm{const} \right] \\
\Delta &= s^2{}_s I + H^T H \\
Z &= \Delta^{-1} H^T Y
\end{aligned}
\tag{4}
$$

using a Restricted Maximum-Likelihood (ReML) algorithm [62]. This process also yields the effective connectivity matrix K corresponding to different remembered stimuli. This matrix maximises the mutual information between the remembered cue and the ensemble activity and is thought to describe synaptic efficacy in a neural ensemble that represents a certain stimulus.

Later, we will consider a network comprising two brain areas that contain memories (complex engram). Assuming two connected brain areas 1 and 2, where area 2 is downstream (and its input is weighted by a connectivity matrix W), we can extend the model given by Equations (2) to describe neural dynamics in the network using a system of neural field equations [53,63,64]

bioRxiv preprint doi: https://doi.org/10.1101/2023.02.28.530474; this version posted March 1, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

$$\begin{aligned} \dot{V}^{1m} &= -\tau_{NA} V^{1m} + K^1 * f \circ V^{1m} + U^1 \\ \dot{V}^{2m} &= -\tau_{NA} V^{2m} + K^2 * f \circ V^{2m} + W \ * f \circ V^{1m} + U^2 \end{aligned} \tag{5}$$

We will use this system below. First, we will derive an extension of the deep neural field from Equation (2) to detect interactions between emerging electric fields produced by neural ensembles and the underlying neural activity (ephaptic coupling). This extension is discussed below.

A model of ephaptic coupling

We wanted to extend Equation (2) and include *ephaptic coupling*. This describes direct influences of the brain's electric fields to individual neurons, that is, interactions between E^e and V^m, see e.g. [35,36,50,65,66]. We followed[67] and added an *ephaptic current* $1/r_i \Delta V_0^e$ to the transmembrane potential of the neural ensemble embedded in E^e. Note that $\nabla h \equiv \partial h / \partial x$ and $\Delta h \equiv \partial^2 h / \partial x^2$, for an arbitrary function *h*. Ephaptic effects change V^m by some fraction *s* of the extracellular potential (EP) V^e. We call *s* the *ephaptic coupling weight*. Mathematically, we describe ephaptic effects by replacing V^m by $V^m + sV_0^e$ in Equation (2). This corresponds to adding a current flowing along the axis of the neural ensemble as the result of top down influence from E^e. The dynamics of V^m can then be described by a neural field equation of the form

$$\dot{V}^m = -\tau_{NA} V^m + K * f \circ (V^m + sV_0^e) + U \tag{6}$$

where V_0^e is the boundary value of the EP on the membrane. We call this model the *ephaptic neural field model*. For *s=0*, Equation (6) becomes Equation (2) that we and others have used

bioRxiv preprint doi: https://doi.org/10.1101/2023.02.28.530474; this version posted March 1, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

before. Equation (6) suggests that V^m changes as a result of a simple decay and presynaptic inputs from other parts of the ensemble, some exogenous, stochastic input U and perturbations due to the extracellular potential V^e that are commensurate to its value on the exterior part of the membrane V_0^e, with constant of proportionality s.

Principal axes

We saw above that the principal axes H of the deep neural field (2) are given by the spatial derivatives of the neural activity. These contain temporal information, that is, fluctuations around baseline activity at different spatial scales. They also provide the dominant frequencies starting from the largest and moving to increasingly smaller spatial scales[51]. The name principal axes derives from training the model (3) with a PCA-like algorithm using the cost function (4). Intuitively, principal axes are like velocity and acceleration when one replaces time with space: velocity describes the distance over which a vehicle has moved in unit time. Then acceleration captures changes in velocity in unit time. Similarly, the second and third principal axes as velocity and acceleration over *space* instead of time. The first principal axis is a linear approximation to fluctuations around the mean and captures activity changes at the largest spatial scale. Then, the second is sensitive to changes of the first axis at a smaller scale, the third is sensitive to changes of the second at an even smaller scale and so on[51].

In case of the ephaptic model (6), the principal axes are the spatial derivatives of the emerging neural activity (including the ephaptic effect) $\hat{X} = V^m + sV_0^e$

$$G_k = \frac{\partial(V^m + sV_0^e)}{\partial x^{(k)}} \qquad (7)$$

where x parameterizes the location along the intracellular fiber axis (Figure 1C). This definition follows from our earlier work (cf. definition of ephaptic axes in [51])

bioRxiv preprint doi: https://doi.org/10.1101/2023.02.28.530474; this version posted March 1, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

after modifying it to include the ephaptic current, similarly to Equation (6) above. It includes the second derivative of the boundary value of the extracellular potential V^e on the membrane V^m, known as activation function[68]. This describes the influence of the external electric field E^e on the membrane currents (ephaptic effects). Different values of the index k correspond to derivatives of different orders. Thus, G_k, k=1,2,3… are called the axes of order k, that is, first, second, third axis etc.

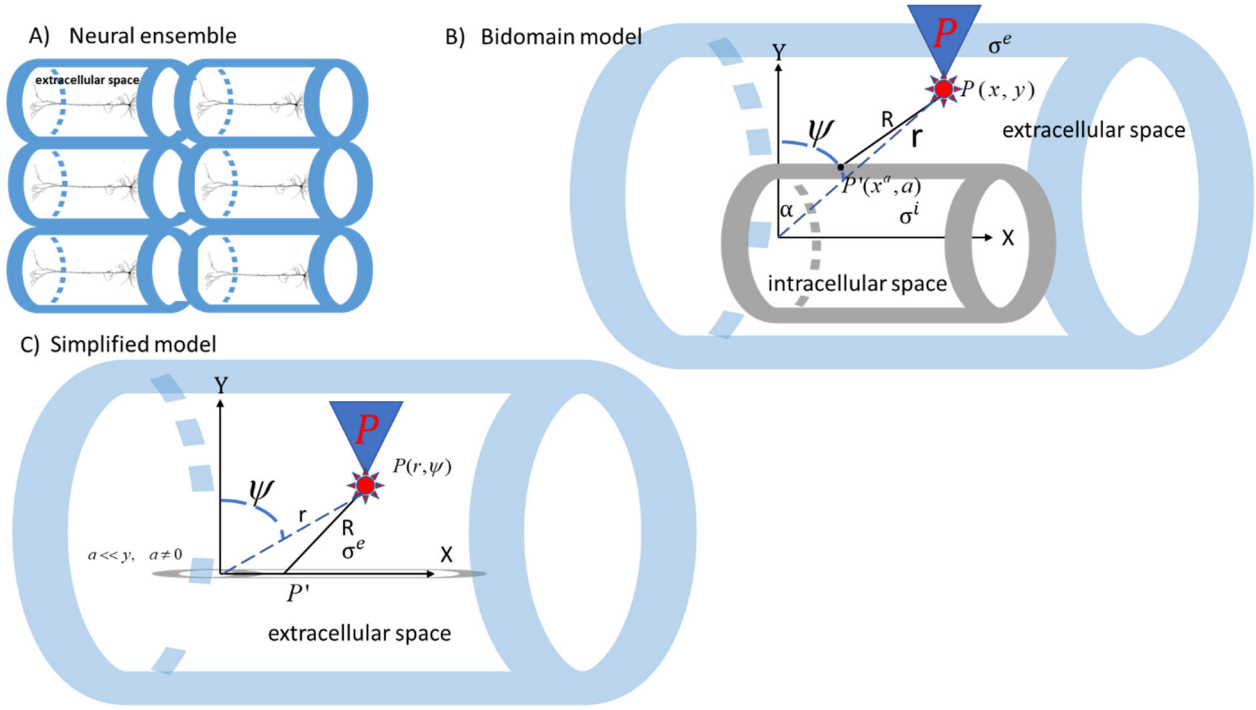

Figure 1.(A) Extracellular space around each neuron within the ensemble (blue cylindrical fibers). (B) Bidomain model for the electric field generated by a cylindrical fiber in a conductor. The extracellular and intracellular space are depicted by blue and grey cylindrical fibers (see Methods for the meaning of various symbols). (C)Simplified bidomain model where the measurement point is located at a vertical distance much larger than the radius of intracellular space.

A model of the ensemble electric field

In [52], we introduced a model of the electric field generated by a neural ensemble. Below, we revisit this model. Neural ensemble activity generates a potential V^e and an electric field $E^e = -\nabla V^e$ in extracellular space. This electric field is the result of the discontinuity in the electric potential $V_0^e - V_0^i$ that gives rise to electric dipole sources and transmembrane currents

bioRxiv preprint doi: https://doi.org/10.1101/2023.02.28.530474; this version posted March 1, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

$1/r_i \Delta V^m$ (V_0^e and V_0^i are the values of the extracellular and intracellular EPs on the two sides of the membrane). Intuitively, E^e is the force needed to move a charged particle, like an ion, between two locations of extracellular space where the potential difference is equal to V^e. The EF is measured by invasive and non invasive electrophysiology techniques like EEG, MEG and LFPs. It acts on charged particles and alters neural activity in a top-down fashion. This is known as ephaptic coupling, to which we will return below.

To describe electric fields arising from the coordinated activity of neural ensembles, we used the *bidomain model* from electromagnetism [69]. This assumes that pyramidal neurons produce an EF parallel to apical dendrites and receive synchronous input. It suggests that the extracellular space of each pyramidal neuron is described by a cylindrical fiber (small blue cylinders in Figure 1A). All these spaces and electric fields are then superimposed and produce the extracellular EF. Superposition suggests that the individual cylindrical fibers of Figure 1A (for each neuron) give rise to the larger fiber surrounding the neural ensemble (light blue cylinder in Figure 1B). It also means that spatial inhomogeneity and asynchronous input will reduce the overall EF. Thus, the electric field predicted by the model provides a lower bound of the real EF. Note that this does not change qualitive results, like ephaptic coupling discussed below as the extracellular and intracellular spaces can be split into smaller parts (cylindrical fibers) where symmetry and synchrony still apply. Because of the cylindrical symmetry, the EP only depends on two spatial variables (x, y), not three. According to the bidomain model, the extracellular EP V^e at a point $P(x,y)$ in the extracellular space (e.g. the location of the LFP electrode denoted by a star in Figure 1B), can be expressed in terms of the Fourier Transform $\widehat{V}^m$ of the transmembrane potential V^m by the following expression, see also [52] for more details,

$$V^e(x,y) = -(4\pi\sigma^e/\sigma^i)FT^{-1}[\widehat{V}^m(k)W(k)]$$

$$W(k) = \frac{I_1(|k|a)K_0(|k|y)}{I_0(|k|a)K_1(|k|a) + \sigma^i/\sigma^e I_1(|k|a)K_0(|k|a)} \qquad (8)$$

bioRxiv preprint doi: https://doi.org/10.1101/2023.02.28.530474; this version posted March 1, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Here, σ^l, $l = \{e, i\}$, are the extra-and intra-cellular space conductivities and $I_0(y), I_1(y)$, a is the radius of the intracellular fiber and $K_0(y), K_1(y)$ are modified Bessel functions of the first and second kind [70]. This equation describes the resting state value of EP during memory delay that the neural ensemble maintains a stimulus.

A mathematical proof of ephaptic coupling and the stability of the electric field

In the next two sections of the *Methods*, we include some mathematical proofs of hypotheses tested in *Results*. Above, we summarized a model of the electric field generated by neural ensembles. In [52], we used this model to compute the EF corresponding to neural ensembles maintaining different memory representations. We found that EFs were more stable than neural activity and contained relatively more information. We suggested that this stability allows the brain to control the latent variables that give rise to the same memory. In other words, we hypothesized that EFs can sculpt and herd neural activity and can act as "guard rails" that funnel the higher dimensional variable neural activity along stable lower-dimensional routes.

Below we provide a mathematical proof of the above hypothesis: that bioelectric fields guide neural activity. In the Results section, we test this hypothesis, using data from a spatial delayed saccade task.

We were interested in interactions between variables expressed at different spatial and temporal scales: bioelectric fields and neural activity. Thus, we used a theory that can systematically describe interactions underlying spontaneous pattern formation in biological and physical systems known as *synergetics* [44,55]. Synergetics studies how individual parts –in our case neurons— produce structures; here, memory representations. It suggests that a biological system, like a neural ensemble, is constrained by so called *control parameters* that impose limitations. When control parameters change, the structures change. A simple example of a control parameter is temperature. When it changes, the state of water molecules can change from solid, to fluid, to air. In synergetics language, the individual elements of the system, e.g. molecules, are called enslaved parts. This is because they are controlled by

bioRxiv preprint doi: https://doi.org/10.1101/2023.02.28.530474; this version posted March 1, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

temperature. Besides control parameters and enslaved parts, synergetics also considers order parameters, that is, low dimensional descriptions of collective dynamics, like the average transmembrane potential V^m that we studied here or other latent variables [71,72] like effective connectivity components [51]. A crucial distinction between control and order parameters is how fast they evolve. When there is a perturbation, like new input to a brain area, the order parameters and enslaved parts evolve fast and the control parameters slowly. Control parameters are very stable compared to order parameters. To put it differently, synergetics suggests a temporal hierarchy comprising, slow control parameters, like temperature or energy [73], faster order parameters and very fast enslaved parts (e.g. oscillations/spiking[25]).

Below, we will use the theory of synergetics to provide a mathematical formulation of ephaptic coupling, that is, the interactions between the ensemble EF , E^e , and the average transmembrane potential V^m. We will present some theoretical arguments that motivate the hypothesis that a slow EF E^e acts as a control parameter that enslaves faster neural activity V^m. In Results, we will test this hypothesis and ask whether ephaptic coupling can be detected in *in vivo* neural data. Finally, our analysis below will also motivate a subsequent hypothesis: do stable electric fields allow the storage of memory engrams in multiple brain areas? This will also be tested in Results.

To describe extracellular field – transmembrane potential E^e - V^m interactions, our starting point is equations that express one quantity in terms of the other, that is E^e in terms of V^m and vice versa. These are Equations (6) and (8): the evolution of transmembrane potential V^m in terms of the extracellular EP V^e is given by the ephaptic model (6). Also, V^e in terms of transmembrane potential V^m is given by the bidomain model (8). For our mathematical (pen and paper) arguments, we need algebraic Equations (without the inverse Fourier transform FT^{-1}). Thus, in Supplementary Methods we show how we can rewrite Equation (8) as a differential algebraic equation. For simplicity, we assume that the LFP electrode is at a large distance compared to the size of the neural ensemble: the radius a of the intracellular fiber (grey) is very small compared to the vertical distance y to the location of the LFP electrode, $a << y$, *cf.* squashed grey cylinder in Figure 1C.

bioRxiv preprint doi: https://doi.org/10.1101/2023.02.28.530474; this version posted March 1, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

From trial to trial the remembered stimulus changes. Thus the EP and the corresponding EF also change, see [52] for details. Assuming a fixed point attractor (steady state), Equation (8) can be written as (see Supplementary Material for details)

$$\begin{aligned} \dot{V}^e &= -\tau_{EP} V^e + \gamma(1/r - L^2/12r^3 + L^4/80r^5)\Delta V^m \\ E^e &= -\nabla V^e \end{aligned} \qquad (9)$$

where $1/\tau_{EP}$ is the rate with which V^e decays to its resting value V_S^e.

Equation (9) expresses the dynamics of the extracellular EP V^e in terms of the transmembrane potential V^m. To describe interactions between these potentials and the corresponding electric fields, we then applied the *slaving principle* from synergetics[44]. This predicts that control parameters evolve more slowly and constrain order parameters and enslaved parts. Examples of the general slaving principle can be found in physics and biology [43]. Haken and colleagues have shown that varying the temperature (control parameter) of a fluid heated from below, various spatial patterns of fluid molecules occur. Also, that attention can be thought of as control variable in multi-stable perception[73,74].

During working memory delay, the slaving principle leads to ephaptic coupling: it predicts that extracellular EP, V^e, enslaves neural activity described by the transmembrane potential V^m. To confirm this, consider the following expansion of V^m and V^e in terms of Fourier series $\begin{pmatrix} V^e \\ V^m \end{pmatrix} = \sum_n \begin{pmatrix} \xi_n \\ \psi_n \end{pmatrix} e^{inx}$. Then, substituting these expansions into Equations (6) and (9), we obtain evolution equations for the Fourier coefficients or modes

$$\begin{aligned} \dot{\xi}_n &= -\tau_{EP}\xi_n + \gamma(1/r - L^2/12r^3 + L^4/80r^5)n^2\psi_n \\ \dot{\psi}_n &= -\tau_{NA}\psi_n + \delta\sum_q \hat{K}_{nq}(\psi_q + s\xi_{q0}) + U \\ \hat{K}_{nq} &= \int K e^{-i2\pi(nx+qy)}dxdy \end{aligned} \qquad (10)$$

bioRxiv preprint doi: https://doi.org/10.1101/2023.02.28.530474; this version posted March 1, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

ξ_n and ψ_n are called the Fourier coefficients or modes of the extracellular potential and neural activity. Below, we call them modes. ξ_{q0} are the values of the extracellular EP on the exterior of the ensemble membrane (surface of grey cylinder in Figure 1B). Intuitively, a Fourier expansion implies that V^m and V^e are superpositions of planar waves e^{inx} with amplitudes given by ξ_n and ψ_n.

We have replaced Equations (6) and (9) that describe the coupling between the extracellular potential and neural activity, V^e and V^m, by Equations (10) that describe the same coupling in terms of modes. Note that in the second equation (10) that the rate of change of neural activity modes, $\dot{\psi}_n$, depends on values of the extracellular potential modes ξ_{q0} on the exterior of the membrane and exogenous stochastic input *U*.

We can now apply the slaving principle of synergetics. This suggests that the instantaneous values of fast relaxing quantities, like the transmembrane potential modes ψ_n, depend on slowly varying quantities, like the extracellular potential coefficients ξ_{q0} above, which slave them [44]. This is ephaptic coupling formulated in the language of synergetics. The transmembrane potential obeys the extracellular potential and the corresponding electric field. In [43,44], several Equations similar to (10) are presented in the context of physics and biology and similar coupling between fast and slow quantities is discussed.

The justification of ξ_{q0} being slow and ψ_n being fast goes as follows: In each trial, the maintained memory (stimulus) changes, thus the input to the memory network from other brain as areas changes. A common assumption in bio-electromagnetism is that the EF is quasi-static, that is, the tissue impedance on top of resistance (reactance) is negligible and electromagnetic propagation effects can be ignored [75]. In [52], we confirmed this and found that the electric field was more stable than neural activity. We found that correlations of single trial estimates of electric fields were higher than correlations of similar neural activity estimates. Since the EF is stable, the Fourier coefficients ξ_{q0} of the extracellular EP V^e vary slowly compared to the Fourier coefficients ψ_n of the transmembrane potential V^m. The damping constant for the extracellular potential is much smaller than the damping constant for neural activity $\tau_{NA} >> \tau_{EP}$. This is also an example of an adiabatic approximation in

bioRxiv preprint doi: https://doi.org/10.1101/2023.02.28.530474; this version posted March 1, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

physics [44].

The fact that neural activity is enslaved by the EP and electric field suggests that its instantaneous values are given in terms of *instantaneous values of the slower fields*. During delay, transmembrane potential is assumed to be in equilibrium, thus $|\dot{\psi}_n| = 0$. Then, Equations (10) yield these instantaneous values of neural activity determined by emerging fields. One can express ψ_n in terms of ξ_n

$$\psi_n = \frac{\delta}{\tau_{NA}} \sum_q \hat{K}_{nq} \left(\psi_q + s\xi_{q0} \right) \tag{11}$$

This equation describes how the fast modes of neural activity are enslaved (driven) by the slow, *stable* modes of the electric field.

To sum, the slaving principle from synergetics predicts that stable electric fields enslave neural activity. They determine the instantaneous values of the transmembrane potential. Mathematically, the slaving principle distinguishes between stable and unstable quantities, like the modes ξ_n and ψ_n . It suggests that the evolution of fast unstable modes is determined by stable modes. This is also related to critical slowing where some modes are strongly correlated over time, e.g. [76–78], see also [52].

A mathematical proof that ephaptic coupling leads to engram complexes

The distinction between stable and unstable modes can be obtained using a mathematical theory known as linear stability analysis. Linear stability analysis of neural network models is often used to express brain responses in terms of key anatomical and biophysical parameters, e.g. [53,55,64,79]. It can also be extended to include nonlinear terms, see [44,74]. Here, we use linear stability analysis to motivate a hypothesis about engram storage in memory networks that will be tested in Results: that ephaptic coupling leads to engram complexes. If ephaptic coupling occurs in a brain area and this exchanges memory information with other brain areas

bioRxiv preprint doi: https://doi.org/10.1101/2023.02.28.530474; this version posted March 1, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

then ephaptic coupling will occur in those areas too. Below, we present a mathematical proof of this hypothesis for two areas. Generalization to an arbitrary number of areas can be done by induction.

Consider two neural ensembles in brain areas (1) FEF and (2) SEF. Dynamics of ensemble activity are given by a system of neural fields of the form of Equations (5). In the linear stability regime, we can assume that the transmembrane potential V^{im} of each ensemble (identified by the upper index j =1,2) includes perturbations in the form of planar waves around baseline V^{io}, that is an equation of the form $V^{jm} \sim V^{jo} + e^{\beta t + ikx}$, [63,79]. Also, similarly to Equation (6) above, ephaptic coupling suggests V^{jm} depends on EP V_0^{je} (its boundary value at the membrane exterior) via the following expressions:

$$\begin{aligned} \dot{V}^{1m} &= -\tau_{NA} V^{1m} + K^1 * f \circ (V^{1m} + s^1 V_0^{1e}) + U^1 \\ \dot{V}^{2m} &= -\tau_{NA} V^{2m} + K^2 * f \circ (V^{2m} + s^2 V_0^{2e}) + W \ * f \circ V^{1m} + U^2 \end{aligned} \tag{12}$$

Here, W is the feedforward connectivity matrix whose entries are weights that scale downstream input to SEF from FEF [57,58]. For mathematical convenience, we consider a vector of extracellular and transmembrane potential functions for the two areas

$$\Phi = [V^{1e}, V^{1m}, V^{2e}, V^{2m}]^T = \sum_n \begin{pmatrix} \xi_n^1 \\ \psi_n^1 \\ \xi_n^2 \\ \psi_n^2 \end{pmatrix} e^{inx} \tag{13}$$

Upper indices denote the area and lower indices the mode order. In the previous section we saw that the slaving principle suggests that the slow, stable field modes ξ_n^1 and ξ_n^2 will constrain ψ_n^1 and ψ_n^2. The order of the expansion (13), n ,(how many modes are needed to faithfully represent the dynamics) can be found using a model fitting procedure (e.g.

bioRxiv preprint doi: https://doi.org/10.1101/2023.02.28.530474; this version posted March 1, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

maximum likelihood or similar) using real data. We will consider this elsewhere. Since we here focus on mathematical arguments, for simplicity, we assume that the first two modes explain most of the observed variance, that is, we keep terms up to 2nd order in Equation (13) (n=1,2)

$$\Phi \approx [\xi_1^1, \psi_1^1]^T e^{ix} + [\xi_2^1, \psi_2^1,]^T e^{2ix} + [\xi_1^2, \psi_1^2]^T e^{ix} + [\xi_2^2, \psi_2^2]^T e^{2ix} \qquad (14)$$

Substituting the above expression in Equations (12) and using the first of Equations (10), we obtain a system of equations

$$\dot{\Phi} = M\Phi + \text{nonlinear terms} \qquad (15)$$

where the matrix M can be expressed in terms of 4x4 matrices A,B,C and D, $M = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$ defined in the *Supplementary Material.* Further, the matrix D can be written as $D = \begin{bmatrix} E & F \\ G & H \end{bmatrix}$ in terms of *2x2* matrices E,F,G and H also included in the Supplementary Material. These matrices include constants and the connectivity matrices K and W appearing in Equation (12). Equation (15) is a linearized system that describes the coupling of extracellular and transmembrane potentials in the two areas in terms of connectivity. Mathematically, for the system to have a solution, that is, for the modes to exist, the determinant of the matrix M needs to be different than zero, $\det(M) \neq 0$. By applying the identity $\det(M) = \det(A - BD^{-1}C)\det(D)$ [70], we obtain

$$\det(M) = \det(A)\det(E - FH^{-1}G)\det(H) \qquad (16)$$

Thus, the condition $\det(M) \neq 0$, requires that $\det(H) \neq 0$ and $\det(A) \neq 0$; the determinants of matrices M, A and H should be non-zero. H is defined by

bioRxiv preprint doi: https://doi.org/10.1101/2023.02.28.530474; this version posted March 1, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

$$H = \begin{bmatrix} \beta + \tau_{EP} & 4Z \\ s\delta \widehat{K}_{22}^2 & \beta + \tau_{NA} + \delta \widehat{K}_{22}^2 \end{bmatrix} \qquad (17)$$
$$Z = \gamma(1/r - L^2/12r^3 + L^4/80r^5)$$

In other words, H is the matrix of coefficients in a (linearized) system of equations describing the coupling between the second extracellular and membrane potential modes in the second region:

$$\begin{aligned} \dot{\xi}_2^2 &= -\tau_{EP}\xi_2^2 + \gamma(1/r - L^2/12r^3 + L^4/80r^5)4\psi_2^2 \\ \dot{\psi}_2^2 &= -\tau_{NA}\psi_2^2 + \delta \widehat{K}_{22}^2(\psi_2^2 + s\xi_{20}^2) \end{aligned} \qquad (18)$$

Then, the condition $\det(H) \neq 0$ implies that the above system has a solution, i.e. the modes exist. In the previous section, we found that using synergetics, Equations (18) imply that the second extracellular potential mode is coupled to neural activity and ephaptic coupling occurs. This is a consequence of ephaptic coupling. Similarly, the condition $\det(A) \neq 0$ means that there is ephaptic coupling in the first area (i.e. first and second modes are coupled).

Crucially, the constraint given by Equation (16) says that assuming that modes exist in both areas ,i.e. the determinant of the matrix M is non-zero, and there is ephaptic coupling in the second area (i.e. the determinant of matrix H is non-zero), then the determinant of matrix A will also be non-zero. Or the other way around. Thus, assuming that ephaptic coupling occurs in one area, then ephaptic coupling will also occur in the other area. By induction, we can show the same result for an arbitrary number of areas. To sum, using a simple mathematical condition (that a linearized system has a solution and the determinant of its coefficient matrix is non-zero), we have concluded that ephaptic coupling will occur in all brain areas that form a memory network or engram complex.

bioRxiv preprint doi: https://doi.org/10.1101/2023.02.28.530474; this version posted March 1, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Granger causality

To test for information transfer between different spatial scales (emerging electric fields and neural activity) and brain areas (FEF and SEF), we used Granger causality (GC; [80,81]) . GC quantifies how the history (past samples) of variable A improve prediction of unknown samples (future samples) of a different variable B. It is based on generalized variances or log likelihood ratios that quantify whether a regression model including variable A fits future samples of variable B better than the restricted regression model based on variable B samples only [82]. Following [48],we evaluated GC as follows: we first used model based VAR modeling to calculate regression coefficients from our data, similar to a discrete stationary vector stochastic process. First, one determines an appropriate order of a VAR model using an information criterion or cross validation[83]. Then, a log-likelihood ratio $F_{A\to B}$ of residual covariance matrices is computed. This corresponds to the full and restricted VAR models and quantifies the *GC strength*, that is, whether the prediction of future values of the variable *B* improves significantly after including past values of *A*. This can be computed using Granger's *F*-test for univariate problems or a chi-square test for a large number of variables [80,81]. GC is often used for the analysis of time series (samples are obtained using measurements at different moments in time). Here, we used GC after considering spatial samples, that is, we obtained measurements at different locations in the neural ensemble and extracellular space. This is discussed further in *Results*.

Representation Similarity Analysis

We used Representation Similarity Analysis[49] (RSA) to assess the similarity of information representation across different brain areas. RSA uses Dissimilarity Matrices (DMs) to summarize how stimulus information is represented by brain responses. Following [60], we built DMs based on time correlations that are thought to underlie working memory representations [84,85]. Each DM entry contained the dissimilarity between trials corresponding to different remembered cued locations. Thus, DMs describe pairwise differences in patterns of neural activity corresponding to different stimuli. To understand whether similar information (cued location) was encoded in different brain areas, we computed the dissimilarity between brain DMs. Following[49] the dissimilarity between dissimilarity matrices, known as deviation, was the correlation distance (1- Spearman correlation; Spearman was used as it does not require a linear correspondence between these matrices

bioRxiv preprint doi: https://doi.org/10.1101/2023.02.28.530474; this version posted March 1, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

contrary to Pearson correlation). Deviations between DMs quantify matches between representation content of brain responses [86]. They measure the correlation distance between each DM and quantify differences of differences: How different are the corresponding pairwise differences in neural activity or electric fields. After calculating deviations of DM matrices one can assess significant correspondence between information stored in different brain areas [87,88].

Results

Ephaptic coupling in in—vivo memory delay data

We first asked whether we could find evidence for ephaptic coupling in our data. We examined *in vivo* LFPs acquired from FEF and SEF during delay in a spatial WM task[10,42]. In[52], we analysed the same data from FEF only. Here we extended our analyses to the FEF-SEF memory network (engram complex). In that earlier work [52], we also introduced a model predicting the electric field generated by a neural ensemble. This yielded a mapping from membrane depolarization to extracellular potential. Depolarization was described by a neural field model trained as an autoencoder using the same LFP data, that we call *deep* neural field, see [41,51]. The term "deep" reflects the bottleneck architecture of the Restricted Maximum-Likelihood (ReML) algorithm used for training. Connectivity in the deep neural field predicting the activity in a neural ensemble maximizes the information between the cue and its representation[51]. Here, we extended that earlier neural field model to include ephaptic coupling, that is the change in the activity of neural ensembles maintaining memory representations as a result of top down influences from the electric field [67] (*Methods*).

Our analyses used two deep neural field models: one model with and another without ephaptic coupling. We call these, the ephaptic and non-ephaptic model. We constructed the ephaptic model by modifying our earlier deep neural field model. Having obtained the activity using the deep neural field model, we then reconstructed the electric field by applying the bidomain model from electromagnetics[89]. Then, following [67,68], we added a term

bioRxiv preprint doi: https://doi.org/10.1101/2023.02.28.530474; this version posted March 1, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

involving the second spatial derivative of the electric field in the original model. This is known as activating function and describes the current density induced by the electric field along the axis of the neural ensemble (denoted by X in Figure 1). Thus, we obtained a model capturing ephaptic effects (*Methods*).

Using our model, we simulated ephaptic effects. These are found by subtracting the oscillatory responses predicted by the non-ephaptic model from the corresponding responses predicted by the ephaptic model. To compute these changes, we considered the oscillatory responses during memory delay. These include fluctuations of neural activity around baseline because of endogenous noise driving the neural ensemble. These fluctuations can be thought of as transient non-Turing patterns (patterns that decay back to baseline) whose Lyapunov exponents determine the frequencies observed in sampled LFP activity. In our model, oscillatory responses are predicted by the principal axes, see also[51] . These are matrix-valued functions of dimensionality $N_T x T$, where N_T is the number of trials and T is the length of the raw LFP time series. Principal axes are obtained after training the model using a PCA-like autoencoder algorithm *(Methods).* We call each entry in these matrices the axis strength. This corresponds to an instantaneous scale factor with which the corresponding component strength must be multiplied to reconstruct the observed LFP. Across-trial averages of principal axes for FEF and SEF were shown in Figures 2 and 3 of [51]. In brief, to understand oscillatory responses during memory maintenance in the presence and absence of ephaptic coupling, we computed the principal axes of the ephaptic and non-ephaptic models.

To assess the effect of the ephaptic coupling on neural activity, we subtracted the original (non-ephaptic) first axis from the corresponding ephaptic model axis and averaged over trials[1]. The results of our analyses are included in Figure 2. Figure 2A (left) shows the relative percent changes due to the ephaptic coupling for FEF. Similarly, Figure 2B (right) shows the corresponding relative changes for SEF. There are six panels in each Figure, each corresponding to a different cued location (angle). This is shown in bottom right of each panel, e.g. the top left panel corresponds to cued location $\theta = 0$ degrees. The vertical panel axes show the relative change in principal axis strength with respect to the original principal axis, after including ephaptic coupling.

[1] The first axis describes fluctuations of neural activity around baseline while higher order axes describe its temporal derivatives, see [51] for a discussion.

bioRxiv preprint doi: https://doi.org/10.1101/2023.02.28.530474; this version posted March 1, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

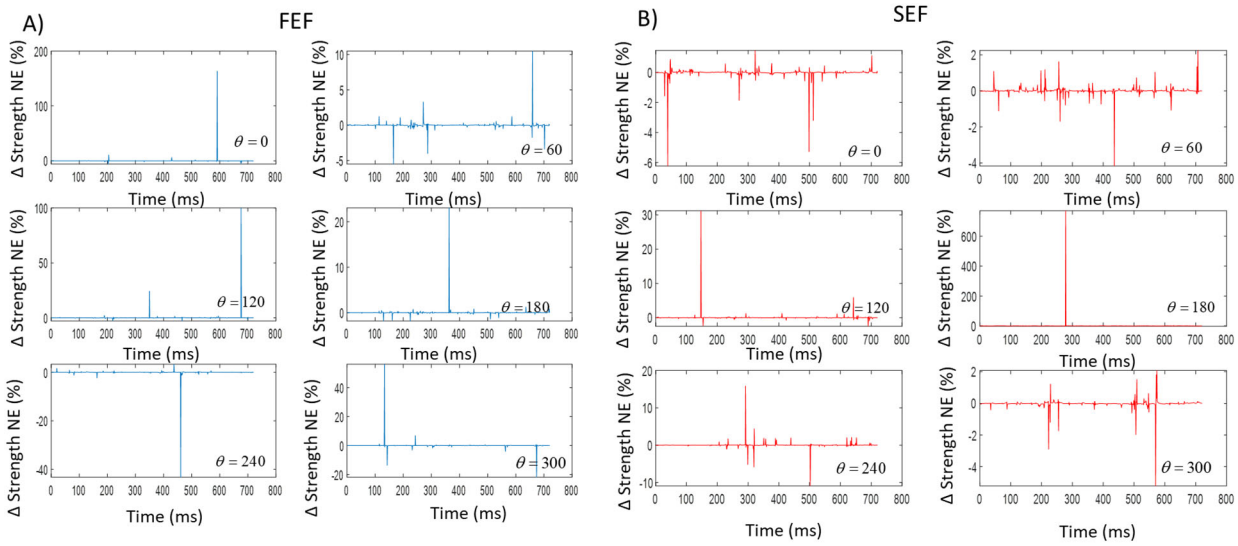

Figure 2.

Relative percent changes in endogenous neural activity predicted by the model (principal axis) due to ephaptic coupling for (A) FEF and (B) SEF. A positive relative change of α% implies that endogenous fluctuations are α times larger than the corresponding fluctuations without the ephaptic coupling and similarly for negative values.

We call these relative changes ephaptic effects. A positive relative change of $\alpha\%$ implies that the principal axis strength is α times larger than the corresponding value without the ephaptic coupling and similarly for negative values. The horizontal panel axes show time in *ms*. During each trial, changes because of ephaptic coupling vary between zero and some maximum value, positive or negative. In general, for both brain areas, maximum changes across stimuli can be split in small and large: for FEF, the locations $\theta = 60, 180, 240$ and 300 degrees, showed maximum changes between 10-40% (we call them *small* effects), while changes for $\theta = 0$ and 120 degrees are one order of magnitude larger, between 100-200% (we call them *large* effects). Similarly, for SEF, maximum changes for locations $\theta = 0, 60$ and 300 degrees were small (between 2-6%), while maximum changes for locations $\theta = 120, 180$ and 240 degrees were large (between 20-600%). Changes in Figure 2 describe how and when the transient wave patterns change because of ephaptic coupling. For example, for cued location at $\theta = 0$ degrees, FEF activity changes by over 160% just before t=600ms (top left panel in Figure 2A). Similarly, SEF activity for $\theta = 0$ is reduced by 5% at t=500ms due to the

bioRxiv preprint doi: https://doi.org/10.1101/2023.02.28.530474; this version posted March 1, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

ephaptic coupling (top left panel in Figure 2B). The results in the remaining panels can be interpreted in a similar way.

To look for evidence of ephaptic coupling, we fitted the ephaptic and non-ephaptic models to LFP data and evaluated goodness of their fits. If the fit of the ephaptic model was better, this would provide evidence of ephaptic coupling. We used previously unseen data for model fitting (data not used for training). Using the ReML algorithm for inference[90], we obtained Free Energy approximations of model evidence. Free Energy is a cost function borrowed from autoencoders that we used to measure goodness of fit. Inference used single trial data and the principal axes as input to infer connectivity, similar to Dynamic Causal Modeling (DCM) and other model fitting approaches [91–93].

We fitted the ephaptic and non-ephaptic models to LFP data from FEF and SEF. We used Bayesian Model Comparison [90,94,95] to find the model that fit the data best [96]. We compared the evidence (how well a model could explain the data) of the two models, the ephaptic and the non-ephaptic. Evidence was computed using a Free Energy aapproximation to model evidence. We fitted the models and found the one that explains data better; this was the model with a (log) Bayes factor (BF) higher than 3 [46]. BF can be thought of as a probabilistic analogue of the odds ratio used in frequentist statistics. This corresponds to a posterior probability of 95% for the winning model. Here, BF describes how likely is the ephaptic model to have generated the sampled LFPs[2]. We assumed that the same model was true for all trials (fixed effects) and asked whether the ephaptic model would fit LFP data better.

BF results are shown in Figure 3A (vertical axis). These are averaged over trials for each cued location. The horizontal axis shows the 6 different locations (angles) cued to hold in working memory. Blue bars denote the BF after fitting FEF data, while red bars after fitting SEF data. A positive BF implies that the non-ephaptic model was more likely; a negative BF that the ephaptic model was. The arrow at the right hand side of Figure 3A facing upwards includes the letters NE = non-ephaptic, while the downwards facing arrow, the letter E=ephaptic wins. BF bars pointing “downwards” provides evidence of ephaptic coupling.

[2] Variance explained by the deep neural field model (non ephaptic) was about 40%, see Supplementary Figure 2A.

bioRxiv preprint doi: https://doi.org/10.1101/2023.02.28.530474; this version posted March 1, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

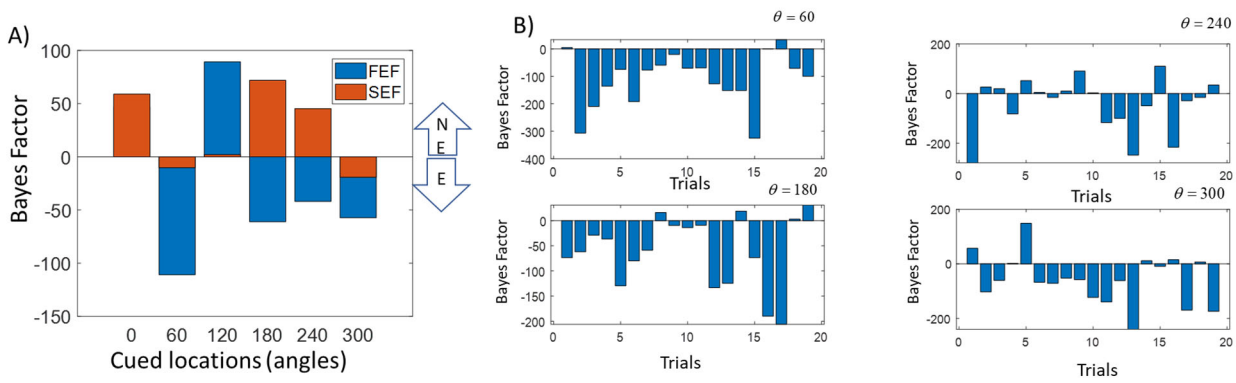

Figure 3. (A) Bayes factor (BF) for different cued locations (horizontal axis). Blue bars denote the BF after fitting FEF data, while red bars after fitting SEF data. A positive BF implies that the non-ephaptic model was more likely; a negative BF that the ephaptic model was. BF bars pointing "downwards" provides evidence of ephaptic coupling, denoted by the E inside the lower arrow. NE in the upper arrow stands for "non-ephaptic". (B)Bayes factor for individual trials and specific cued angles. Different trials are shown on the horizontal axis. The corresponding cued angles are shown at the top right corner of each plot. The ephaptic model fits the data better for most trials.

Using model comparison, we found that in FEF, the ephaptic model was more likely for cued locations at $\theta = 60, 180, 240$ and 300 degrees (BF=*-120,70,45* and *55* respectively; blue bars in Figure 3A). To make sure the ephaptic model fitted single trial data better, Figure 3B shows the BF for individual trials for $\theta = 60, 180, 240$ and 300 degrees, i.e. when the ephaptic model was more likely in FEF. We confirmed that the ephaptic model was better in most trials. BF estimates are between BF=*20-310* for $\theta = 60$ degrees, BF=*10-200* for $\theta = 180$ and $\theta = 240$ degrees, and BF=*5-220* for $\theta = 300$ degrees. SEF results are similar (orange bars in Figure 3A). The ephaptic model was more likely for $\theta = 60$ and 300 degrees (BF=*-10* and *20* respectively)[3].

Although results were robust over trials, we did not find evidence of ephaptic coupling across all cued locations. Comparing results in Figures 3A and 2, we concluded that the ephaptic model explained the LFP data better only when ephaptic effects were small.

When the model predicted large ephaptic effects (large LFP variability), fitting the model to

[3] As expected the complexity of both models was very similar, see Supplementary Figure 2C which shows the difference in complexity between models. All estimates are between 0-0.6, which is less than 0.5% of the BF factor shown in Figure 3A. Note also that for $\theta = 0$ degrees the non ephaptic model was more likely.

bioRxiv preprint doi: https://doi.org/10.1101/2023.02.28.530474; this version posted March 1, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

data did not confirm ephaptic coupling. Thus, either there was no ephaptic coupling or the model overfitted. The first explanation is refuted by the results of the next section. The second explanation is consistent with them and also follows from model assumptions. The model is valid for linear, i.e. small fluctuations (fractions of fluctuations of membrane potential around baseline) with a smooth spatial profile[41]. It does not hold for large ephaptic effects (multiple times the average fluctuations).

Below, we used an alternative approach, Granger causality to bypass the linearity assumption above and test for *nonlinear* interactions between the electric fields and neural activity. Crucially, this approach also allowed us to obtain the directionality of these interactions. This is discussed below.

Top down information transfer from emerging electric fields to neuronal ensembles

Above, we found that, when endogenous fluctuations were small, a model in which neural ensemble activity is coupled to the electric field (ephaptic model) explained the LFP data better than a model without ephaptic coupling. We next tested for ephaptic coupling more generally, during large endogenous fluctuations. To do so, we used Granger Causality (GC, see *Methods*). GC is a data-driven method for determining the directionality of information flow between stochastic variables[81]. Crucially, GC also provides the directionality of the interactions between the electric field and neural activity. In other words, GC allows us to test whether the electric field guides neural activity or the other way around. In [41], we suggested that electric fields can act as "guard rails" that funnel the higher dimensional variable neural activity along stable lower- dimensional routes. We tested this hypothesis directly using GC.

In its common use, GC is applied to time series data and assesses whether knowing the past of one variable (A) helps predict the future of another variable (B) better than just using the past of B alone. If so, one concludes that information flows from variable A to B. Flow is thought to occur over time, similarly to the flow of a water molecule that flows in a river. In

bioRxiv preprint doi: https://doi.org/10.1101/2023.02.28.530474; this version posted March 1, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

neuroscience, GC is used to describe how information flows in the brain, using sampled time series from different areas[48].

One way to compute GC is by first calculating the covariance function, that is, how strongly a time series is related to itself or another time series. This requires p samples, that is, measurements at p time steps earlier or later[82]. Implicit in this calculation, there is an assumption of finite p, or, that the information flows at a *finite speed* from the variable A to B. Here, we focused on the information flow between the electric field and neural activity (i.e., electric field and neural activity are the variables A and B). It is well known that interactions involving the electric field transfer information very close to the speed of light, which is practically infinite. Thus, the assumption of finite velocity in GC analyses does not hold here. If the field is coupled to neural activity, information contained in it also contains information about neural activity at each point of the cortical patch occupied by the neural ensemble at the *same time*. This is similar to applications in geophysics where GC and recordings of the earth's gravitational field are used to e.g. find what kind of minerals exist deep below the surface[97]. In other words, the emerging electric field contains instantaneous information about neural activity in the same way that gravitational field contains instantaneous information about the masses of minerals underneath. We here used this idea from geophysics after replacing the gravitational with the electric field and mineral masses with neural activity.

Because of the practically infinite speed of information propagation, we followed a slightly unusual GC analysis where we replaced time with space samples. We considered snapshots of time series and computed the GC over space. We used data from a single time point. Data included the spatial profiles of neural activity and contemporaneous electric field snapshots. At each time point, we asked whether knowing the electric field helps predict the value of neural activity in a neighboring location, where activity had not been measured yet, better than using recordings of neural activity alone. GC measures interactions in both directions, thus our analyses answered the reverse question too: whether knowing neural activity helps predict the electric field. Our analyses are summarized in Figure 4.

bioRxiv preprint doi: https://doi.org/10.1101/2023.02.28.530474; this version posted March 1, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

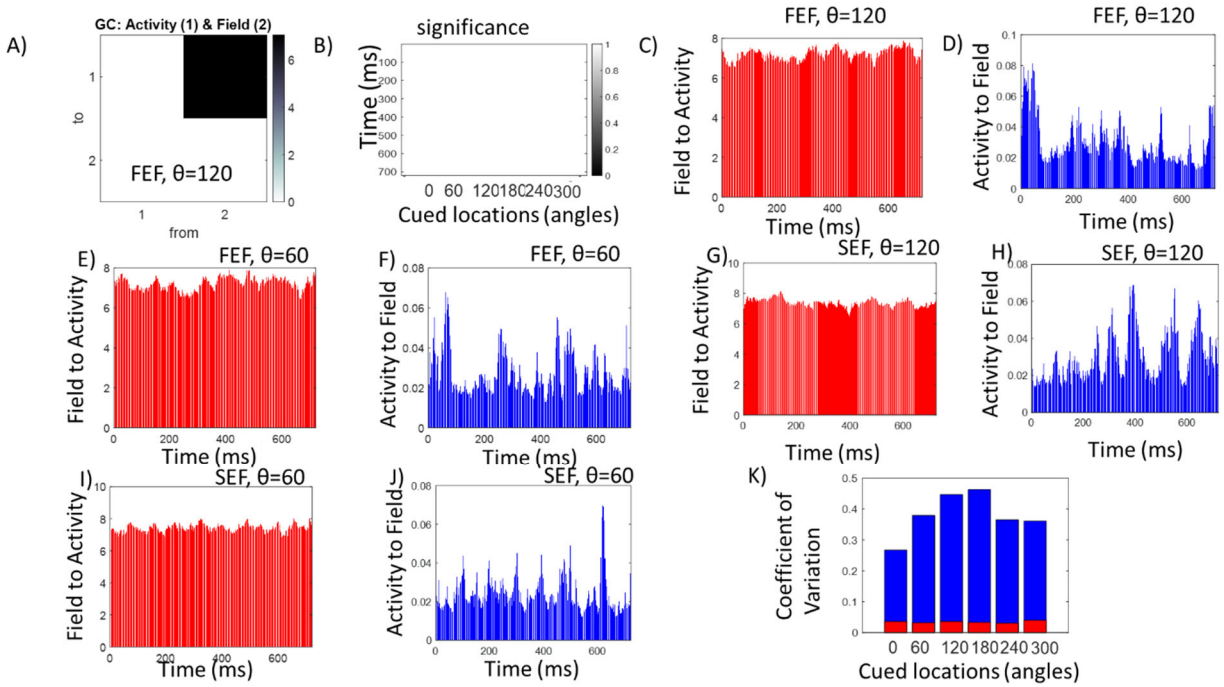

Figure 4. (A) Granger causality (GC) strengths for field-to-neural activity interactions. (B)Significance of GC strengths in (A). (C—F) Examples of individual GC strengths corresponding to each time point during delay for cued locations at $\theta = 120$ *and* $\theta = 60$ *degrees, computed using FEF data. (G – J) Similar to C—F above, for SEF data. (K) Coefficients of variation for GC strengths (vertical axis) for all remembered cued locations (horizontal axis) computed using FEF data. Red bars depict variability in field-to-activity GC strengths and blue bars depict variability in activity-to-field GC strengths.*

Following[48], we used an *F*-test to assess CG strength *(Methods)*. First, using LFPs from FEF we calculated GC strength and averaged across all time points. Results are shown in Figure 4A for $\theta = 120$ degrees. The top right quadrant (from field to activity) has a GC strength of GC=*7.83*, while the bottom left (from activity to field) has a GC strength of GC=*0.04*. Results for other angles are very similar (not shown). Figure 4B shows *F*-test significance in the field to activity direction for all cued locations. Time points are shown on the vertical axis and cued locations on the horizontal. White entries correspond to a significant GC strength.

We found that field-to-activity GC was significant across all time points and for all remembered angles. Examples of individual GC strengths corresponding to each time point during delay for $\theta = 120$ and $\theta = 60$ degrees, are shown in Figures 5C – F for FEF and Figures 5G – J for SEF. GC strengths are shown on the vertical axis and time points on the horizontal. Field to activity GC strengths are shown in red, while activity to field GC

bioRxiv preprint doi: https://doi.org/10.1101/2023.02.28.530474; this version posted March 1, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

strengths in blue. In FEF, field to activity GC strengths range between GC=*6.42–7.86* ($\theta = 120$ degrees) and GC=*6.41–7.98* ($\theta = 60$ degrees). Activity to field GC strengths range between GC=*0.01–0.08* ($\theta = 120$ degrees) and GC=*0.01–0.06* ($\theta = 60$ degrees). Results for SEF were very similar: field to activity GC strengths range between GC=*6.45–8.19* ($\theta = 120$ degrees) and GC=*6.76–8.04* ($\theta = 60$ degrees). Activity to field GC strengths range between GC=*0.01–0.07* ($\theta = 120$ degrees) and GC=*0.01–0.08* ($\theta = 60$ degrees).

All in all, the above results suggest that across all remembered cued locations, GC was much larger in the field to activity than the reverse direction in both FEF and SEF. This confirms our earlier results about in vivo ephaptic coupling in memory ensembles using BMC and extends them for all stimuli. The electric field drives the neural activity. It funnels the high dimensional varying neural activity along stable lower dimensional routes – as suggested in[41].

Another result from [41] was that electric fields were more stable than neural activity. This was confirmed here using GC analysis. Comparing red and blue bars in Figures 5C-J (both FEF and SEF results), we observed that activity-to-field GC strengths varied more over time than field-to-activity GC strengths. This difference in temporal variability between electric field and neural activity is formally assessed using coefficients of variation (CV). Figure 4K shows the CVs for GC strengths (vertical axis) for all remembered cued locations (horizontal axis) using FEF recordings. Red bars depict variability in field-to-activity GC strengths and blue bars depict variability in activity-to-field GC strengths. We found that variability was much higher in the activity-to-field direction. Blue bars corresponding to different cued locations were much larger (CV=*28–47%*) than red bars (CV=*2–4%*). Results for SEF were similar (not shown).

Ephaptic coupling and the stability of the electric field found here using coefficients of variation based on Granger Causality follows also from the theory of synergetics. The theory suggests that order parameters, like the electric field, affect enslaved parts, like neural activity (ephaptic coupling). Synergetics also suggests that control parameters (fields) are also more stable than enslaved parts (neural activity). See *Methods* for a mathematical, i.e. pen and paper, proof of this result.

bioRxiv preprint doi: https://doi.org/10.1101/2023.02.28.530474; this version posted March 1, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Electric fields guide information transfer in engram complexes

Next, we considered causal interactions between electric fields and neural activity in engram complexes across cortical areas. Recall that such complexes include brain areas that maintain memories[23] connected via mono- or poly-synaptic connections. We examined engrams formed by FEF and SEF in our spatial delay task. We studied information transfer between these brain areas using GC. The analyses below are like those in the previous section. The difference is that below, variables A and B are electric fields (or neural activity) from *different* brain areas, as opposed to the same areas considered above.

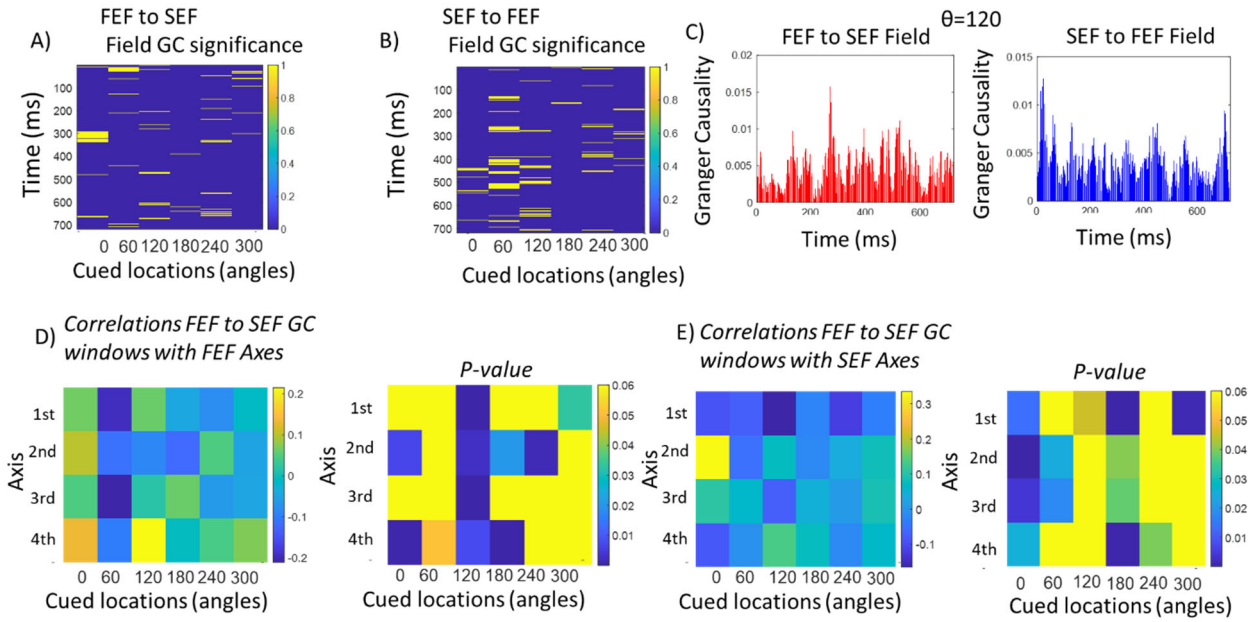

Figure 5.(A) Time points of significant GC field interactions from FEF to SEF for all cued locations. Time is shown on the vertical axis and cued locations on the horizontal. Significant interactions are shown in yellow. (B)Similar to 5A. Significant GC field interactions for the reverse direction, from SEF to FEF. (C)GC strengths (vertical axis) of FEF to SEF (left panel) and SEF to FEF (right panel) field interactions across time (horizontal axis) for $\theta = 120$ degrees. (D)Correlations (left panel) and p values (right panel) between FEF principal axes and temporal windows during which GC field interactions from FEF to SEF were significant. Principal axes are shown on the vertical axis (from first to fourth as we move downwards) and cued locations on the horizontal axis.(E) Similar to Figure 5D for SEF principal axes.

We first computed the GC strength based on electric fields in FEF and SEF. This is shown in Figure 5. The corresponding results based on neural activity are shown in Supplementary Figure 1. We first considered at which exact time points interactions between the two areas

bioRxiv preprint doi: https://doi.org/10.1101/2023.02.28.530474; this version posted March 1, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

were significant. These time points are shown in Figures 5A and 5B. Significant interactions are shown in yellow for all cued locations in the FEF to SEF direction in Figure 5A. Remembered cued locations are shown on the horizontal axis while time points on the vertical. Figure 5B has the same format as 5A and shows the corresponding results in the opposite, SEF to FEF direction. For example, for $\theta = 0$ degrees, significant electric field interactions in the FEF to SEF directions were observed at sparse intervals between times *t= 290–310ms* and around *t=690ms* (yellow lines in the first column of Figure 5A). In the SEF to FEF direction, such interactions were found around *t=310, 540,490, 540* and *680ms* (Figure 5B).

Example field-to-field GC strengths for a cued location at $\theta = 120$ degrees are shown in Figure 5C. FEF to SEF field GC strengths are shown in the left panel (red). GC strength in the reverse direction is shown in the right panel (blue). GC strengths are on the vertical axis. Time points are on the horizontal axis. Strengths have similar ranges in both directions during the delay period. We found similar results using neural activity (Supplementary Figure 1). Like the results based on electric fields discussed above, Supplementary Figures 1A and 1B reveal temporal windows of information transfer between FEF and SEF at the neural activity level, confirming our results using electric fields. Supplementary Figure 1C shows GC strengths in both directions. Interactions at the level of neural activity are expected: we found above that electric fields guide neural activity and that there were significant interactions between FEF and SEF electric fields. Thus, information transfer at the field level (Figures 5A and 5B) leads to information transfer at the neural activity level. Clearly, GC interactions at the level of neural activity are sparser than the corresponding GC strengths at the electric field level and this is replicated across all cued angles (results not shown). There are fewer red and blue lines in the left and right panels of Supplementary Figure 1C compared to Figure 5C. At several time points, GC strengths based on neural activity were zero, while GC strengths based on fields were not. This confirms the stability and robustness of the electric field found above and in our earlier work.

Are the temporal windows during which significant field interactions occur related to neural activity fluctuations? If so, this would mean that the dynamics (fluctuations) of neural ensembles in FEF and SEF are linked to the information transfer between them. This is what we tested next. We asked whether the temporal profile of significant field interactions found using GC above (yellow lines in Figures 5A and 5B) follows the neural dynamics in each

bioRxiv preprint doi: https://doi.org/10.1101/2023.02.28.530474; this version posted March 1, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

brain area. Our hypothesis was that significant field interactions would occur while neural activity fluctuations were relatively large. We thus looked for correlations between the temporal windows (epochs) during which GC significant field interactions took place and principal axes. Recall that principal axes provide the fluctuations of neural activity around baseline at different spatial scales[51]. We thus computed the correlations between the first, second, third and fourth principal axes and the epochs during which field GC was significant.

In Figure 5D, we show correlations (left panel) and the corresponding *p* values (right panel) between FEF principal axes and temporal windows during which electric field GC interactions from FEF to SEF were significant. Principal axes are shown on the vertical axis (from first to fourth as we move downwards) and cued locations on the horizontal axis. Figure 5E includes the corresponding results for SEF principal axes. Different colors in the *p* values (right) panel correspond to different significance levels –where we have lumped together all *p* values above the significance threshold (p=0.05) and shown them in yellow. The same visualization is followed in Figure 5E and Supplementary Figures 1D and 1E. In brief, yellow entries denote non- significant correlations.

Overall, for both FEF and SEF and all cued angles, the temporal windows during which FEF to SEF CG strengths based on electric fields were significant, correlated with principal axes, i.e. endogenous fluctuations around baseline. P-values in each column (cued location) in the right panels in Figures 5D and 5E includes non-yellow, i.e. significant correlations. Interestingly, this was not the case for GC strengths based on neural activity. For certain angles, there were no significant correlations between GC strengths based on neural activity and fluctuations (principal axes). This was the case for correlations with FEF axes for $\theta = 0$ degrees (right panel in Supplementary Figure 1D) and with SEF axes for $\theta = 180$ and 240 degrees (right panel in Supplementary Figures 1E).

Thus, we found that fluctuations around baseline activity in both areas correlated with the temporal windows of significant field GC interactions. The evolution of information transfer between FEF and SEF follows the dynamics of the neural ensembles in these areas. The link between information transfer (significant GC interaction windows) and neural dynamics appears stronger at the level of electric fields. This suggests that fields are more stable than neural activity.

bioRxiv preprint doi: https://doi.org/10.1101/2023.02.28.530474; this version posted March 1, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Above we found significant interactions at the level of electric fields in both directions between FEF and SEF (Figures 5A and 5B). We also found that interactions in the FEF to SEF direction followed the dynamics of neural ensembles (Figures 5D and 5E). Interestingly, interactions for several cued locations GC strengths in the reverse direction were non-significant. Supplementary Figure 2B (left panel) shows this was the case for FEF fluctuations and cued locations at $\theta = 180, 240$ and 300 degrees. The right panel in the same figure shows absence of significant correlations with SEF fluctuations (SEF axes) for $\theta = 120, 180$ and 240 degrees. This suggests that information flow in the memory network seems to follow FEF, not SEF neural ensemble activity. SEF activity at the same time, includes both information flowing out from SEF and reverberating delay activity in the FEF—SEF network.

The same memory is stored by electric fields in different brain areas

In the previous section, we used Granger causality and found that information was transferred between brain areas, FEF and SEF, during memory maintenance. Our hypothesis was that data were recorded from sites that are part of engram complexes.

To confirm this, we asked whether representations (engrams) in each site corresponded to the same memory. To test for similarity between information content we used *Representation Similarity Analysis* (*RSA*[49], *Methods*). First, one constructs Dissimilarity Matrices (DMs) based on correlation distance to evaluate the similarity between memory representations. DMs describe pairwise differences in patterns of neural activity or electric fields corresponding to different cued locations. In turn, correlation distances between DMs, known as deviations, express second order differences, that is, differences in pairwise differences in neural activity or electric fields in different brain areas for the same cued locations. We used deviations to test for significant correspondence between memory representations [60,87,88].

bioRxiv preprint doi: https://doi.org/10.1101/2023.02.28.530474; this version posted March 1, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

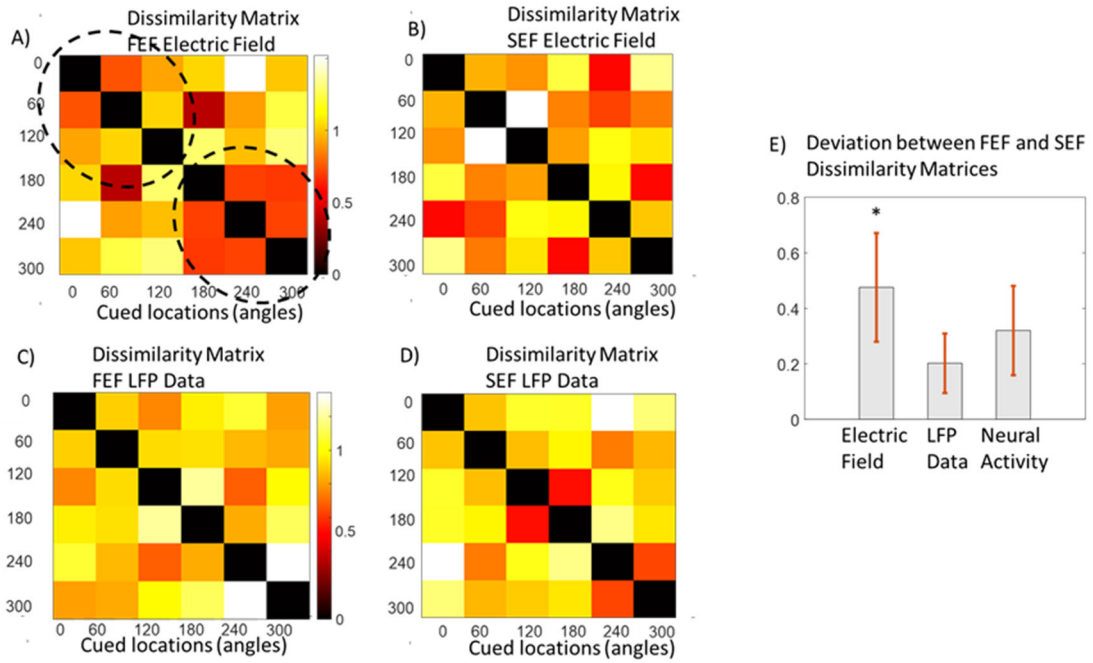

Figure 6. (A) Representation Dissimilarity Matrix (RDM) computed using FEF electric fields. Notice the lattice structure shown inside the dashed ellipses reminiscent of topographic clustering in FEF. (B) RDM computed using SEF electric fields. (C) RDM computed using FEF data. (D) RDM computed using SEF data. (E) Deviations (second order correlations) between RDMs. Deviation for electric field RDMs was the only that was significant (denoted by an asterisk above the leftmost bar; significance at the p<0.05 level). Error bars denote the standard errors (N=100).

We first constructed DMs for FEF and SEF based on three different sets of data: electric fields, LFPs and neural activity. Fields and activity were reconstructed using our model *(Methods)*. Our results are shown in Figure 6 and Supplementary Figure 3. Figures 6A and 6B include the DMs for FEF and SEF electric fields respectively. Figures 6C and 6D include the corresponding RDMs based on LFPs and Supplementary Figures 3A and 3B include RDMs based on neural activity. Different colors correspond to different dissimilarities (1-correlation) for each of the six possible cued locations.

Correlations were computed between trials corresponding to the same stimulus for all possible stimulus pairs after averaging over time. The higher the dissimilarity the more variability in the way information is represented. In other words, DMs illustrate the geometry of stimulus space, that is, how different cued locations are distributed into the space spanned by the activity of the underlying neural ensemble or its electric field. This provides a visualization of how dynamics in different brain areas represent memories. It can reveal clusters implying categorical representations or smooth variations along stimulus dimensions

bioRxiv preprint doi: https://doi.org/10.1101/2023.02.28.530474; this version posted March 1, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

that link to behavior. The overall structure of matrices in Figures 6A and 6B describes how the electric field representations differ between pairs of cued angles. Diagonal terms have zero dissimilarity as expected. Representations were different between stimuli (red and yellow entries, $d \geq .4$). This is also the case for other RDMs in Figures 6C and 6D as well as Supplementary Figures 3A and 3B. Interestingly, FEF RDMs based on electric field and neural activity Figure 6A and Supplementary Figure 3A show a lattice structure: representations corresponding to the upper ($\theta = 0, 60$ and 120 degrees) and lower ($\theta = 180, 240$ and 300 degrees) hemifield form distinct clusters, shown by ellipses. This is reminiscent of topographic clustering in FEF, that is known to contain topographically organized responses and visual map[98,99]. It is also in accord with a similar organization of functional and effective connectivity found using the same dataset in [10].

To confirm that representations contained the same memories, we then computed the deviations between DMs[49,60]. Our results are shown in Figure 6E. Deviation is a second order correlation distance, that is, the distance between correlation distances shown in DMs. It allows us to quantify matches between memory representations in the two areas. The smaller the deviation the closer the match. To test whether two DMs were related, we used fixed effects randomization test. We simulated the null distribution by reordering rows (10,000 relabelings) and obtained a distribution of correlations (the null hypothesis is that the two DMs were unrelated). If the actual correlation we had obtained fall within the top 5% of the simulated null distribution, then we reject the null hypothesis: the two DMs are related. Figure 6E shows that the deviation for electric fields was larger ($d \approx .5$) than that computed using neural activity ($d \approx .3$) which, in turn, is larger than the deviation computed using LFPs ($d \approx .2$). Crucially, the randomization test reveals that only the DMs based on electric fields are significantly related (denoted by an asterisk above the leftmost bar in Figure 6E; significant deviations at the $p<0.05$ level). Error bars denote the standard errors. They depict the variability of deviations (had we chosen different stimuli from the same population; $N=100$, see [49]).

To sum, we found significantly related dissimilarity matrices in FEF and SEF computed using electric fields, but not LFPs or reconstructed neural activity. This suggests that memory representations in the two areas, known as engrams, are linked at the electric field level. Crucially, these similarities in memory representations across two areas were not apparent in

bioRxiv preprint doi: https://doi.org/10.1101/2023.02.28.530474; this version posted March 1, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

LFP recordings. Taken together with our earlier result that emerging electric fields seem to guide information transfer, our result here suggests that electric fields mediate the transfer of memories and their latent states between brain areas. Ephaptic interactions occur in areas where engrams are found. See *Methods* for a mathematical proof of this result.

Discussion

We found evidence for in vivo ephaptic coupling from two cortical areas, the FEF and SEF, during performance of a spatial delayed response task. We found that ephaptic coupling from bioelectrical fields is causative, it influences neural activity, sculpting and guiding it to form engram complexes. We found that, in each brain area, information was transferred from bioelectric fields to neurons. Also, stable, robust fields allowed for memory transfer between FEF and SEF engrams. Neural activity appeared to contain less information and was more variable. In short, like a conductor of an orchestra, where neurons are the musicians, the bioelectric field influences each neuron and orchestrates the engram, the symphony.

To demonstrate ephaptic effects, we used biophysical modelling and Granger causality. We used a model that can describe neural ensemble connectivity, synaptic filtering, and electric fields. In previous work, we estimated the effective connectivity in neural ensembles and their electric fields[10,41]. We found that electric fields carry stimulus information, are robust and can act as "guardrails" that stabilize and funnel the underlying neural activity. We showed that fields were more stable than neural activity and could be used to decode remembered cued locations better.

Here, we used the same model and tested whether including ephaptic effects resulted in better fits to LFP data. The model was used for both learning and inference. It first learned the connectivity parameters. These were subsequently used as priors to reconstruct single trial neural activity and bioelectric field estimates. This revealed ephaptic effects when endogenous fluctuations were small, as expected from the linearity assumption of our model. Granger Causality (GC) applied to time snapshots confirmed these ephaptic effects during large endogenous fluctuations and also allowed us to determine directionality.

bioRxiv preprint doi: https://doi.org/10.1101/2023.02.28.530474; this version posted March 1, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Our results were consistent with the communication through coherence hypothesis (CTC [30]). According to CTC, neural ensembles synchronize in a way that creates bursts of excitation and inhibition and allows information to propagate from one area to the other during certain temporal windows. We took CTC one step further to suggest that this communication is guided by emerging electric fields. First, we found that between area GC strengths based on fields were larger than the corresponding estimates based on neural activity. Second, for each brain areal GC strengths were much larger in the field-to-activity than in the reverse direction. Third, the temporal windows during which FEF to SEF interactions take place followed the dynamics of neural ensembles in these areas. Taken together, the above results suggest that electric fields guide information transfer between areas.

The electric fields were more stable than neural activity, i.e., had less representational drift. This concurs with earlier results where electric field estimates were more often correlated across trials, i.e. more stable, compared to neural activity estimates[41]. Here, we found similar results. The coefficients of variation associated with field -to-activity GC strengths were smaller than the corresponding coefficients based on activity-to-field interactions. Also, GC strengths of interactions between FEF and SEF neural activity were sparser over time than the corresponding strengths based on electric fields.

Using Representation Similarity Analysis (RSA[49,87]), we also confirmed that electric fields emerging from FEF and SEF ensembles contained the same information. RSA assesses matches between memory representations in different brain areas. Information can be represented at different levels, e.g. in neural activity or electric fields. We found that FEF and SEF representations contained similar information only when we used electric field data for RSA analysis – not LFP or neural activity. Thus, memory representations seem to be linked at the electric field level.

Overall, our results suggest that in addition to synaptic transmission, information transfer might be guided in a top-down fashion by electric fields. In mathematical language, electric fields are a control parameter. This term appears in the theory of synergetics from complex systems [44,74] . Examples of control parameters include energy [43,74], and feedback attention signals in a binocular rivalry task [73]. A control parameter has two features that the electric field has: it is stable and evolves at a slower time scale than enslaved parts (i.e. neural activity). In *Methods*, using mathematical arguments, we explained these features and showed

bioRxiv preprint doi: https://doi.org/10.1101/2023.02.28.530474; this version posted March 1, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

how ephaptic coupling follows from the slaving principle[43] (see also discussion below). We also showed that if ephaptic coupling occurs in one brain area in a memory network (engram complex) it will occur in all other brain areas.

The idea that electrical fields play a role in neural ensemble formation has a long history. The connection between memories, connectivity and electric fields was noted early. The term engram complex was coined by German biologist Richard Semon, who, over a century ago, suggested that memories are stored in groups of neurons in multiple brain areas[100]. Then, according to Semon's law of ecphory, memory recall happens when an appropriate electric field is generated —an energetic "condition" similar to memory registration is achieved during recall [12,13,100].

The importance of the electric field has also been emphasized in recent synaptic plasticity studies. These have revealed that learning and memory change scaffolding proteins that regulate synaptic functions, like trafficking and binding of NMDA or other receptors [101]. In turn, protein changes result in changes of synaptic activity and of the electric field in the extracellular space. Thus, synaptic activity is not dictated solely by electrical elements, the receptors, charged particles and currents, but also chemical elements, like scaffold proteins. Both electrical and chemical elements determine the electric field in the extracellular space[2]. Receptors occupy synapses with some probability, and can vary from trial to trial where the same memory is recalled. This also means that different neurons form ensembles in different trials where the same memory is maintained, a phenomenon known as representational drift[102].

It is now known that the brain's endogenous electric field feeds back to the activity of individual ion channels and alters their neuronal firing, i.e., there is ephaptic coupling[35], see [39] for a review. The pioneering study by Eccles and Jaeger[103] showed ephaptic effects on ion currents in synaptic cleft. McFadden and other authors have taken the importance of ephaptic coupling one step further: they have linked it to conscious awareness and hypothesized that it can be used for computation that occurs momentarily and is distributed over space [104–107]. Direct evidence of ephaptic coupling has been found in slices[37–39] . Testing such hypotheses and in vivo ephaptic effects in general is more difficult. Electrodes are far from the neural ensemble and multiple groups of neurons are activated at the same time. Further, chemical processes like electrodiffusion and others alter the electric fields[108].

bioRxiv preprint doi: https://doi.org/10.1101/2023.02.28.530474; this version posted March 1, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Here, we used a variety of computational techniques to provide in vivo evidence.

The low-dimensional stability of electric fields can help the brain with memory maintenance and cognitive processing in general. Synergetics suggests that latent states, like connectivity, can be reliably transferred between brain areas, in accord with modern engram theory [17]. This is orchestrated by control parameters. In synergetics, latent states are called order parameters[71,72]. The theory posits that order and control parameters exist in all self-organized dynamical systems (e.g. molecules, fluids) and therefore the brain. They emerge because of self-organization and capture collective dynamics of a large part of the system's individual parts. Importantly, parts, order and control parameters evolve at different timescales that are separate: control (bioelectric fields, slowest), order parameters (e.g. effective connectivity, oscillation frequency, intermediate) and enslaved parts (spiking, fastest).

This separation of timescales follows from the center manifold theorem. Haken[109] pointed out this separation is crucial for consciousness. Order parameters evolve slowly and this "can be interpreted as a phase transition from subliminal to conscious phase". They sent essential information to other brain areas. This is like Mooney faces associated with gamma oscillations and conscious experience[110]. The order parameter is the frequency of oscillations. This hypothesis resonates with modern engram theories that suggest that memories are the result of coordinated activity within and across ensembles[18]. Order parameters control the spiking of a large number of neurons. In [96], we showed that during a working memory task, when the cognitive capacity limit was exceeded, synchrony between oscillatory responses in PFC, FEF and LIP broke down and the monkey made errors. That is, order parameters were different when the monkey could vs. when he could not remember.

Synergetics suggests that control parameters guide order parameters and constrain enslaved parts. Neurons give rise to the ensemble and this, in turn, determines the function of each neuron through ephaptic coupling. This is an application of the slaving principle. This is also a difference between synergetics and dimensionality reduction approaches[111,112]. Like dimensionality reduction, synergetics uses latent states. But it also uses control parameters. These evolve at an even slower timescale than latent states and spiking and are characteristic for each state of the brain, e.g. each memory. Synergetics suggests that control parameters are somehow fixed in the sense that when they change, the brain goes to a different stable state, similar to phase transitions in thermodynamics [113].

bioRxiv preprint doi: https://doi.org/10.1101/2023.02.28.530474; this version posted March 1, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

In sum, using biophysical modeling, machine learning and Granger causality we provided some evidence supporting the hypothesis that bioelectric fields are a control variable that enslaves neural activity. This can have implications for modern BCI, where electric field manipulations are used to control neurons so that activity reverts to a healthy state and patient behavior is abolished.

References

1. James, W. (1890). The principles of. Psychology *2*, 94.

2. Queenan, B.N., Ryan, T.J., Gazzaniga, M.S., and Gallistel, C.R. (2017). On the research of time past: the hunt for the substrate of memory. Annals of the New York Academy of Sciences *1396*, 108–125.

3. Buschman, T.J., Siegel, M., Roy, J.E., and Miller, E.K. (2011). Neural substrates of cognitive capacity limitations. Proceedings of the National Academy of Sciences *108*, 11252–11255.

4. Pinotsis, D.A., and Miller, E.K. (2021). Beyond dimension reduction: Stable electric fields emerge from and allow representational drift. bioRxiv.

5. Fujisawa, S., Amarasingham, A., Harrison, M.T., and Buzsáki, G. (2008). Behavior-dependent short-term assembly dynamics in the medial prefrontal cortex. Nature neuroscience *11*, 823–833.

6. Yuste, R. (2015). From the neuron doctrine to neural networks. Nature reviews neuroscience *16*, 487–497.

7. Tayler, K.K., Tanaka, K.Z., Reijmers, L.G., and Wiltgen, B.J. (2013). Reactivation of neural ensembles during the retrieval of recent and remote memory. Current Biology *23*, 99–106.

8. Pfau, D., Pnevmatikakis, E.A., and Paninski, L. (2013). Robust learning of low-dimensional dynamics from large neural ensembles. Advances in neural information processing systems *26*.

9. Pinotsis, D.A., and Miller, E.K. (2017). New approaches for studying cortical representations. In AAAI Spring Symposium-Technical Report (AAAI), pp. 613–615.

10. Pinotsis, D.A., Brincat, S.L., and Miller, E.K. (2017). On memories, neural ensembles and mental flexibility. NeuroImage *157*, 297–313.

11. Buschman, T.J., Denovellis, E.L., Diogo, C., Bullock, D., and Miller, E.K. (2012).

bioRxiv preprint doi: https://doi.org/10.1101/2023.02.28.530474; this version posted March 1, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Synchronous oscillatory neural ensembles for rules in the prefrontal cortex. Neuron *76*, 838–846.

12. Josselyn, S.A., Köhler, S., and Frankland, P.W. (2015). Finding the engram. Nature Reviews Neuroscience *16*, 521–534.

13. Thompson, R.F. (1976). The search for the engram. American Psychologist *31*, 209.

14. Gordon, J.W., Scangos, G.A., Plotkin, D.J., Barbosa, J.A., and Ruddle, F.H. (1980). Genetic transformation of mouse embryos by microinjection of purified DNA. Proceedings of the National Academy of Sciences *77*, 7380–7384.

15. Guzowski, J.F., Timlin, J.A., Roysam, B., McNaughton, B.L., Worley, P.F., and Barnes, C.A. (2005). Mapping behaviorally relevant neural circuits with immediate-early gene expression. Current opinion in neurobiology *15*, 599–606.

16. Fenno, L., Yizhar, O., and Deisseroth, K. (2011). The development and application of optogenetics. Annual review of neuroscience *34*, 389.

17. Ryan, T.J., Roy, D.S., Pignatelli, M., Arons, A., and Tonegawa, S. (2015). Engram cells retain memory under retrograde amnesia. Science *348*, 1007–1013.

18. Tonegawa, S., Pignatelli, M., Roy, D.S., and Ryan, T.J. (2015). Memory engram storage and retrieval. Current opinion in neurobiology *35*, 101–109.

19. Roy, D.S., Park, Y.-G., Ogawa, S.K., Cho, J.H., Choi, H., Kamensky, L., Martin, J., Chung, K., and Tonegawa, S. (2019). Brain-wide mapping of contextual fear memory engram ensembles supports the dispersed engram complex hypothesis. BioRxiv, 668483.

20. Poo, M., Pignatelli, M., Ryan, T.J., Tonegawa, S., Bonhoeffer, T., Martin, K.C., Rudenko, A., Tsai, L.-H., Tsien, R.W., and Fishell, G. (2016). What is memory? The present state of the engram. BMC biology *14*, 1–18.

21. Squire, L.R., and Alvarez, P. (1995). Retrograde amnesia and memory consolidation: a neurobiological perspective. Current opinion in neurobiology *5*, 169–177.

22. Nadel, L., and Moscovitch, M. (1997). Memory consolidation, retrograde amnesia and the hippocampal complex. Current opinion in neurobiology *7*, 217–227.

23. Tonegawa, S., Liu, X., Ramirez, S., and Redondo, R. (2015). Memory Engram Cells Have Come of Age. Neuron *87*, 918–931. 10.1016/j.neuron.2015.08.002.

24. Lundqvist, M., Herman, P., Warden, M.R., Brincat, S.L., and Miller, E.K. (2018). Gamma and beta bursts during working memory readout suggest roles in its volitional control. Nature communications *9*, 394.

25. Miller, E.K., Lundqvist, M., and Bastos, A.M. (2018). Working Memory 2.0. Neuron *100*, 463–475.

26. Harris, K.D., Csicsvari, J., Hirase, H., Dragoi, G., and Buzsáki, G. (2003). Organization of cell assemblies in the hippocampus. Nature *424*, 552–556.

bioRxiv preprint doi: https://doi.org/10.1101/2023.02.28.530474; this version posted March 1, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

27. Miller, J.K., Ayzenshtat, I., Carrillo-Reid, L., and Yuste, R. (2014). Visual stimuli recruit intrinsically generated cortical ensembles. Proceedings of the National Academy of Sciences *111*, E4053–E4061.

28. Singer, W. (1999). Neuronal synchrony: a versatile code for the definition of relations? Neuron *24*, 49–65.

29. Koch, C. (2004). The quest for consciousness a neurobiological approach.

30. Fries, P. (2015). Rhythms for cognition: communication through coherence. Neuron *88*, 220–235.

31. Reinhart, R.M., and Nguyen, J.A. (2019). Working memory revived in older adults by synchronizing rhythmic brain circuits. Nature neuroscience *22*, 820–827.

32. Lakatos, P., Gross, J., and Thut, G. (2019). A new unifying account of the roles of neuronal entrainment. Current Biology *29*, R890–R905.

33. Engel, A.K., and Singer, W. (2001). Temporal binding and the neural correlates of sensory awareness. Trends in cognitive sciences *5*, 16–25.

34. Moore, J.W., and Obhi, S.S. (2012). Intentional binding and the sense of agency: a review. Consciousness and cognition *21*, 546–561.

35. Anastassiou, C.A., Perin, R., Markram, H., and Koch, C. (2011). Ephaptic coupling of cortical neurons. Nature neuroscience *14*, 217–223.

36. Ruffini, G., Salvador, R., Tadayon, E., Sanchez-Todo, R., Pascual-Leone, A., and Santarnecchi, E. (2020). Realistic modeling of mesoscopic ephaptic coupling in the human brain. PLoS computational biology *16*, e1007923.

37. Chiang, C.-C., Shivacharan, R.S., Wei, X., Gonzalez-Reyes, L.E., and Durand, D.M. (2019). Slow periodic activity in the longitudinal hippocampal slice can self-propagate non-synaptically by a mechanism consistent with ephaptic coupling. The Journal of physiology *597*, 249–269.

38. Jefferys, J.G., de la Prida, L.M., Wendling, F., Bragin, A., Avoli, M., Timofeev, I., and da Silva, F.H.L. (2012). Mechanisms of physiological and epileptic HFO generation. Progress in neurobiology *98*, 250–264.

39. Anastassiou, C.A., and Koch, C. (2015). Ephaptic coupling to endogenous electric field activity: why bother? Current opinion in neurobiology *31*, 95–103.

40. Kitamura, T., Ogawa, S.K., Roy, D.S., Okuyama, T., Morrissey, M.D., Smith, L.M., Redondo, R.L., and Tonegawa, S. (2017). Engrams and circuits crucial for systems consolidation of a memory. Science *356*, 73–78.

41. Pinotsis, D.A., and Miller, E.K. (2022). Beyond dimension reduction: Stable electric fields emerge from and allow representational drift. NeuroImage *253*, 119058.

42. Jia, N., Brincat, S.L., Salazar-Gómez, A.F., Panko, M., Guenther, F.H., and Miller, E.K. (2017). Decoding of intended saccade direction in an oculomotor brain–computer

bioRxiv preprint doi: https://doi.org/10.1101/2023.02.28.530474; this version posted March 1, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

interface. Journal of neural engineering *14*, 046007.

43. Haken, H. (2012). Complex Systems—Operational Approaches in Neurobiology, Physics, and Computers: Proceedings of the International Symposium on Synergetics at Schloß Elmau, Bavaria, May 6–11, 1985 (Springer Science & Business Media).

44. Haken, H. (1987). Thermodynamics—synergetics—life.

45. Haken, H., Kelso, J.S., and Bunz, H. (1985). A theoretical model of phase transitions in human hand movements. Biological cybernetics *51*, 347–356.

46. Kass, R.E., and Raftery, A.E. (1995). Bayes factors. Journal of the American Statistical Association *90*, 773–795.

47. Friston, K., and Penny, W. (2011). Post hoc Bayesian model selection. Neuroimage *56*, 2089–2099.

48. Barnett, L., and Seth, A.K. (2014). The MVGC multivariate Granger causality toolbox: a new approach to Granger-causal inference. Journal of neuroscience methods *223*, 50–68.

49. Kriegeskorte, N., Mur, M., and Bandettini, P.A. (2008). Representational similarity analysis-connecting the branches of systems neuroscience. Frontiers in systems neuroscience *2*, 4.

50. Fröhlich, F., and McCormick, D.A. (2010). Endogenous electric fields may guide neocortical network activity. Neuron *67*, 129–143.

51. Pinotsis, D.A., Brincat, S.L., and Miller, E.K. (2017). On memories, neural ensembles and mental flexibility. NeuroImage *157*, 297–313.

53. Coombes, S. (2005). Waves, bumps, and patterns in neural field theories. Biological Cybernetics *93*, 91–108.

54. Deco, G., Jirsa, V.K., Robinson, P.A., Breakspear, M., and Friston, K. (2008). The Dynamic Brain: From Spiking Neurons to Neural Masses and Cortical Fields. Plos Computational Biology *4*.

55. Jirsa, V.K., and Haken, H. (1996). Field theory of electromagnetic brain activity. Physical Review Letters *77*, 960–963.

56. Robinson, P.A., Zhao, X., Aquino, K.M., Griffiths, J.D., Sarkar, S., and Mehta-Pandejee, G. (2016). Eigenmodes of brain activity: Neural field theory predictions and comparison with experiment. NeuroImage *142*, 79–98.

57. Grossberg, S. (1967). Nonlinear difference-differential equations in prediction and learning theory. Proceedings of the National Academy of Sciences of the United States of America *58*, 1329.

58. Wilson, H.R., and Cowan, J.D. (1973). Mathematical Theory of Functional Dynamics of Cortical and Thalamic Nervous-Tissue. Kybernetik *13*, 55–80.

bioRxiv preprint doi: https://doi.org/10.1101/2023.02.28.530474; this version posted March 1, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

59. LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. nature *521*, 436.

60. Pinotsis, D.A., Siegel, M., and Miller, E.K. (2019). Sensory processing and categorization in cortical and deep neural networks. NeuroImage *202*, 116118.

61. Pinotsis, D.A., Moran, R.J., and Friston, K.J. (2012). Dynamic causal modeling with neural fields. Neuroimage *59*, 1261–1274.

62. Harville, D.A. (1977). Maximum likelihood approaches to variance component estimation and to related problems. Journal of the American statistical association *72*, 320–338.

63. Grindrod, P., and Pinotsis, D.A. (2011). On the spectra of certain integro-differential-delay problems with applications in neurodynamics. Physica D: Nonlinear Phenomena *240*, 13–20.

64. Pinotsis, D.A., Hansen, E., Friston, K.J., and Jirsa, V.K. (2013). Anatomical connectivity and the resting state activity of large cortical networks. NeuroImage *65*, 127–138. http://dx.doi.org/10.1016/j.neuroimage.2012.10.016.

65. Rebollo, B., Telenczuk, B., Navarro-Guzman, A., Destexhe, A., and Sanchez-Vives, M.V. (2021). Modulation of intercolumnar synchronization by endogenous electric fields in cerebral cortex. Science Advances *7*, eabc7772.

66. Schmidt, H., Hahn, G., Deco, G., and Knösche, T.R. (2021). Ephaptic coupling in white matter fibre bundles modulates axonal transmission delays. PLOS Computational Biology *17*, e1007858.

67. Goldwyn, J.H., McLaughlin, M., Verschooten, E., Joris, P.X., and Rinzel, J. (2017). Signatures of somatic inhibition and dendritic excitation in auditory brainstem field potentials. Journal of Neuroscience *37*, 10451–10467.

68. Danner, S.M., Wenger, C., and Rattay, F. (2011). Electrical stimulation of myelinated axons: An interactive tutorial supported by computer simulation. Saarbrücken (VDM 2011).

69. Jackson, J.D. (1999). Classical electrodynamics (American Association of Physics Teachers).

70. Abramowitz, M., Stegun, I.A., and Romer, R.H. (1988). Handbook of mathematical functions with formulas, graphs, and mathematical tables (American Association of Physics Teachers).

71. Gallego, J.A., Perich, M.G., Chowdhury, R.H., Solla, S.A., and Miller, L.E. (2020). Long-term stability of cortical population dynamics underlying consistent behavior. Nature neuroscience *23*, 260–270.

72. Yu, B.M., Cunningham, J.P., Santhanam, G., Ryu, S., Shenoy, K.V., and Sahani, M. (2008). Gaussian-process factor analysis for low-dimensional single-trial analysis of neural population activity. Advances in neural information processing systems *21*, 1881–1888.

73. Ditzinger, T., and Haken, H. (1989). Oscillations in the perception of ambiguous patterns a model based on synergetics. Biological Cybernetics *61*, 279–287.

bioRxiv preprint doi: https://doi.org/10.1101/2023.02.28.530474; this version posted March 1, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

74. Basar, E., Flohr, H., Haken, H., and Mandell, A.J. (2012). Synergetics of the Brain: Proceedings of the International Symposium on Synergetics at Schloß Elmau, Bavaria, May 2–7, 1983 (Springer Science & Business Media).

75. Nunez, P.L. (1998). Macro-neocortical dynamics, cognition, and EEG. Proceedings of the 2nd International Conference on Bioelectromagnetism, 45-46 204.

76. Bassett, D.S., and Bullmore, E.T. (2009). Human brain networks in health and disease. Current opinion in neurology *22*, 340.

77. Chialvo, D.R. (2010). Emergent complex neural dynamics. Nature physics *6*, 744–750.

78. Kitzbichler, M.G., Smith, M.L., Christensen, S.R., and Bullmore, E. (2009). Broadband Criticality of Human Brain Network Synchronization. Plos Computational Biology *5*, e1000314.

79. Pinotsis, D.A., and Friston, K.J. (2010). Neural fields, spectral responses and lateral connections. Neuroimage.

80. Geweke, J. (1982). Measurement of linear dependence and feedback between multiple time series. Journal of the American statistical association *77*, 304–313.

81. Granger, C.W. (1969). Investigating causal relations by econometric models and cross-spectral methods. Econometrica: journal of the Econometric Society, 424–438.

82. Friston, K., Moran, R., and Seth, A.K. (2013). Analysing connectivity with Granger causality and dynamic causal modelling. Current opinion in neurobiology *23*, 172–178.

83. Burnham, K.P., and Anderson, D.R. (1998). Practical use of the information-theoretic approach. In Model selection and inference (Springer), pp. 75–117.

84. Inagaki, H.K., Fontolan, L., Romani, S., and Svoboda, K. (2017). Discrete attractor dynamics underlying selective persistent activity in frontal cortex. biorxiv, 203448.

85. Wallis, G., Stokes, M., Cousijn, H., Woolrich, M., and Nobre, A.C. (2015). Frontoparietal and cingulo-opercular networks play dissociable roles in control of working memory. Journal of Cognitive Neuroscience *27*, 2019–2034.

86. Kriegeskorte, N. (2011). Pattern-information analysis: from stimulus decoding to computational-model testing. Neuroimage *56*, 411–421.

87. Diedrichsen, J., and Kriegeskorte, N. (2017). Representational models: A common framework for understanding encoding, pattern-component, and representational-similarity analysis. PLoS computational biology *13*, e1005508.

88. Peterson, J.C., Abbott, J.T., and Griffiths, T.L. (2018). Evaluating (and improving) the correspondence between deep neural networks and human representations. Cognitive science *42*, 2648–2669.

89. Roth, B.J. (1997). Electrical conductivity values used with the bidomain model of cardiac tissue. IEEE Transactions on Biomedical Engineering *44*, 326–328.

bioRxiv preprint doi: https://doi.org/10.1101/2023.02.28.530474; this version posted March 1, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

90. Friston, K. (2008). Hierarchical Models in the Brain. Plos Computational Biology *4*.

91. Freestone, D.R., Karoly, P.J., Nešić, D., Aram, P., Cook, M.J., and Grayden, D.B. (2014). Estimation of effective connectivity via data-driven neural modeling. Frontiers in neuroscience, 383.

92. Oesterle, J., Behrens, C., Schröder, C., Hermann, T., Euler, T., Franke, K., Smith, R.G., Zeck, G., and Berens, P. (2020). Bayesian inference for biophysical neuron models enables stimulus optimization for retinal neuroprosthetics. Elife *9*.

93. Pinotsis, D.A., Moran, R.J., and Friston, K.J. (2012). Dynamic causal modeling with neural fields. Neuroimage *59*, 1261–1274.

94. Pinotsis, D.A., Brunet, N., Bastos, A., Bosman, C.A., Litvak, V., Fries, P., and Friston, K.J. (2014). Contrast gain-control and horizontal interactions in V1: a DCM study. Neuroimage *92*, 143–155.

95. Friston, K., Mattout, J., Trujillo-Barreto, N., Ashburner, J., and Penny, W. (2007). Variational free energy and the Laplace approximation. Neuroimage *34*, 220–234.

96. Pinotsis, D.A., Buschman, T.J., and Miller, E.K. (2018). Working Memory Load Modulates Neuronal Coupling. Cerebral Cortex.

97. Marques, F., Matos, J.X., Sousa, P., Represas, P., Araújo, V., Carvalho, J., Morais, I., Pacheco, N., Albardeiro, L., and Gonçalves, P. (2019). The role of land gravity data in the Neves-Corvo mine discovery and its use in present-day exploration and new target generation. First Break *37*, 97–102.

98. Funahashi, S., Bruce, C.J., and Goldman-Rakic, P.S. (1989). Mnemonic coding of visual space in the monkey's dorsolateral prefrontal cortex. Journal of neurophysiology *61*, 331–349.

99. Thompson, K.G., and Bichot, N.P. (2005). A visual salience map in the primate frontal eye field. Progress in brain research *147*, 249–262.

100. Semon, R., Duffy, B., and Lee, V. (2018). Mnemic psychology (Routledge).

101. Kim, E., and Sheng, M. (2004). PDZ domain proteins of synapses. Nature Reviews Neuroscience *5*, 771–781.

102. Rule, M.E., O'Leary, T., and Harvey, C.D. (2019). Causes and consequences of representational drift. Current opinion in neurobiology *58*, 141–147.

103. Eccles, J.C., and Jaeger, J.C. (1958). The relationship between the mode of operation and the dimensions of the junctional regions at synapses and motor end-organs. Proceedings of the Royal Society of London. Series B-Biological Sciences *148*, 38–56.

104. Fingelkurts, A.A., Fingelkurts, A.A., and Neves, C.F. (2012). "Machine" consciousness and "artificial" thought: An operational architectonics model guided approach. Brain research *1428*, 80–92.

105. John, E.R. (2005). From synchronous neuronal discharges to subjective awareness?

bioRxiv preprint doi: https://doi.org/10.1101/2023.02.28.530474; this version posted March 1, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Progress in Brain Research *150*, 143–593.

106. McFadden, J. (2020). Integrating information in the brain's EM field: the cemi field theory of consciousness. Neuroscience of Consciousness *2020*, niaa016.

107. Pockett, S. (2000). The nature of consciousness: A hypothesis (IUniverse).

108. Savtchenko, L.P., Poo, M.M., and Rusakov, D.A. (2017). Electrodiffusion phenomena in neuroscience: a neglected companion. Nature reviews Neuroscience *18*, 598–612.

109. Haken, H. (2006). Synergetics of brain function. International journal of psychophysiology *60*, 110–124.

110. Lachaux, J.-P., George, N., Tallon-Baudry, C., Martinerie, J., Hugueville, L., Minotti, L., Kahane, P., and Renault, B. (2005). The many faces of the gamma band response to complex visual stimuli. Neuroimage *25*, 491–501.

111. Gao, P., and Ganguli, S. (2015). On simplicity and complexity in the brave new world of large-scale neuroscience. Current opinion in neurobiology *32*, 148–155.

112. Jazayeri, M., and Ostojic, S. (2021). Interpreting neural computations by examining intrinsic and embedding dimensionality of neural activity. arXiv preprint arXiv:2107.04084.

113. Domb, C. (2000). Phase transitions and critical phenomena (Elsevier).

114. Clark, J., and Plonsey, R. (1968). The extracellular potential field of the single active nerve fiber in a volume conductor. Biophysical journal *8*, 842.

115. Plonsey, R. (1974). The active fiber in a volume conductor. IEEE transactions on biomedical engineering, 371–381.