Arousal as a universal embedding for spatiotemporal brain dynamics

Ryan V. Raut[1,2,*], Zachary P. Rosenthal[3], Xiaodan Wang[4], Hanyang Miao[5], Zhanqi Zhang[6], Jin-Moo Lee[5], Marcus E. Raichle[4,5], Adam Q. Bauer[4], Steven L. Brunton[7], Bingni W. Brunton[8], and J. Nathan Kutz[9]

[1]Allen Institute, Seattle, WA, USA
[2]Department of Physiology & Biophysics, University of Washington, Seattle, WA, USA
[3]Department of Psychiatry, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA
[4]Mallinckrodt Institute of Radiology, Washington University in St. Louis, St. Louis, MO, USA
[5]Department of Neurology, Washington University in St. Louis, St. Louis, MO, USA
[6]Department of Computer Science & Engineering, University of California San Diego, La Jolla, CA, USA
[7]Department of Mechanical Engineering, University of Washington, Seattle, WA, USA
[8]Department of Biology, University of Washington, Seattle, WA, USA
[9]Department of Applied Mathematics, University of Washington, Seattle, WA, USA
[*]Correspondence to: ryan.raut@alleninstitute.org

Abstract

Neural activity in awake organisms shows widespread and spatiotemporally diverse correlations with behavioral and physiological measurements. We propose that this covariation reflects in part the dynamics of a unified, multidimensional arousal-related process that regulates brain-wide physiology on the timescale of seconds. By framing this interpretation within dynamical systems theory, we arrive at a surprising prediction: that a single, scalar measurement of arousal (e.g., pupil diameter) should suffice to reconstruct the continuous evolution of multidimensional, spatiotemporal measurements of large-scale brain physiology. To test this hypothesis, we perform multimodal, cortex-wide optical imaging and behavioral monitoring in awake mice. We demonstrate that spatiotemporal measurements of neuronal calcium, metabolism, and brain blood-oxygen can be accurately and parsimoniously modeled from a low-dimensional state-space reconstructed from the time history of pupil diameter. Extending this framework to behavioral and electrophysiological measurements from the Allen Brain Observatory, we demonstrate the ability to integrate diverse experimental data into a unified generative model via mappings from an intrinsic arousal manifold. Our results support the hypothesis that spontaneous, spatially structured fluctuations in brain-wide physiology—widely interpreted to reflect regionally-specific neural communication—are in large part reflections of an arousal-related process. This enriched view of arousal dynamics has broad implications for interpreting observations of brain, body, and behavior as measured across modalities, contexts, and scales.

Introduction

The past decade has seen a proliferation of research into the organizing principles, physiology, and function of ongoing brain activity and brain "states" as observed across various recording modalities, spatiotemporal scales, and species [1–4]. Ultimately, it is of interest to understand how such wide-ranging phenomena are coordinated within a functioning brain. However, as different domains of neuroscience have evolved around unique subsets of observables, the integration of this research into a unified framework remains an outstanding challenge [5, 6].

Interpreting ongoing neural activity is further complicated by its widespread, spatiotemporally heterogeneous relationships with behavioral and physiological variables [7–12]. For instance, in awake mice, individual cells and brain regions show reliable temporal offsets and multiphasic patterns in relation to spontaneous whisker movements or running bouts [7, 8, 13–15]. Such findings have motivated significant attention toward decomposing and interpreting neural variance that appears intermingled with state

changes and movements [16–18]. Notwithstanding efforts to statistically [7, 9, 11] or mechanistically [19–22] decompose this covariance, there is lingering mystery surrounding the endogenous processes that generate and continuously orchestrate the underlying variation in time [23].

In this work, we propose a parsimonious explanation that ties together these observations: that neural activity, organismal physiology, and behavior are jointly regulated through an intrinsic, arousal-related process that continuously evolves over a latent manifold on the timescale of seconds. This unified perspective stems from the observation that nominally arousal-related fluctuations pervade—and hence, may help to bridge—effectively all domains of (neuro-)physiology and behavior [24–31]. Most recently, ongoing arousal fluctuations—indexed by scalar quantities derived from pupillometry, locomotion, cardiorespiratory physiology, or electroencephalography (EEG) [24, 32]—have emerged as a principal statistical factor of global neural variance [7, 33]. Nonetheless, the interpretation and integration of these findings is hindered by the longstanding challenge of precisely defining what "arousal" actually is [26, 34]. Although reducing this construct to individual biological mechanisms can enhance clarity, it fails to address the organism-wide coordination implicit in arousal [24]. As an alternative approach to demystify this construct, we strengthen and extend the above findings by linking variation in common arousal indices to the dynamics of a specific underlying process. We show that this process—which we refer to as *arousal dynamics*—is inherently multidimensional, and is nonlinearly coupled to observables, ultimately accounting for even greater spatiotemporal heterogeneity in brain measurements than is conventionally attributed to arousal (e.g., based on a single factor in linear regression). Crucially, we ground our explanatory framework in (data-driven representations of) this process and its couplings, rather than a statistical factor or inferred biological mechanisms.

A key motivation for our framework is to provide a unified basis for brain state research pursued across two largely separate paradigms. The first, primarily involving invasive recordings in awake mice, has established the regional dependence of cortical activity (especially at frequencies above 1 Hz) upon ongoing fluctuations in arousal and behavioral state [8, 9, 11, 35]; however, general principles have remained elusive [1, 5]. The second, originating in non-invasive human neuroimaging, has established general principles of spontaneous fluctuations in large-scale brain activity (particularly below 0.1 Hz) [36–38], which are characterized by a topographically organized covariance structure. This structure is widely thought to emerge from regionally specific neuronal interactions [39–41]; as such, arousal is not conventionally understood as fundamental to the physiology or phenomenology of interest. Instead, in this paradigm, arousal enters as either a slow or intermittent modulator of interregional covariance [31, 42–47], or as a distinct (i.e., physiologically and/or linearly separable) global component [48–51]. Yet, growing evidence indicates that global spatiotemporal processes (1) predominate at frequencies below 0.1 Hz, (2) can account for topographic structure [52–57], (3) covary with arousal indices on the timescale of seconds [37, 42, 46, 51, 54, 58], and, crucially, (4) remain dynamically interdependent, regardless of how they are decomposed [56, 59–62]. Together, we believe these themes invite a radical reinterpretation of this paradigm [63]—in particular, one which ultimately aligns this paradigm with the first. In short, we propose that (1) the spatial structure of arousal modulation arises from a complex, nondecomposable spatiotemporal process, and that (2) this process constitutes the primary generative mechanism underlying topographically organized covariance structure.

In this study, we test a central prediction of this unified account: that a scalar index of arousal suffices to accurately reconstruct multidimensional measurements of large-scale spatiotemporal brain dynamics. Inherent in this prediction is the hypothesis that "arousal", as typically operationalized, is underpinned by a specific process that is sufficiently complex—namely, multidimensional [64]—to sustain nonequilibrium behavior, such as persistent fluctuations and traveling waves. To recover the latent dimensions of this hypothesized process, we employ a time delay embedding [65–67]—a powerful technique in data-driven dynamical systems [68] which, under certain conditions, theoretically enables reconstruction of dynamics on a manifold from incomplete, even scalar, measurements. This theory is further strengthened by the Koopman perspective of modern dynamical systems [69], which shows that time delay coordinates provide universal embeddings across a range of observables [70].

Through this framework, we provide experimental support for our predictions across simultaneous wide-field recordings of neuronal calcium, metabolism, and hemodynamics, showing that the spatiotemporal evolution of these measurements is largely reducible to a common latent process that is continuously tracked by pupil diameter. In particular, we show that a time delay embedding of pupil diameter may be used to reconstruct a latent arousal manifold; time-invariant mappings from this manifold predict substantial variation in diverse modalities both within and between mice, suggesting the arousal state-space

as a generalizable embedding for brain and organismal dynamics. These findings reveal a deep connection between the mechanisms supporting brain-wide spatiotemporal organization and organism-wide regulation.

Results

We performed simultaneous multimodal widefield optical imaging [71] and face videography in awake, head-fixed mice. Fluorescence across the dorsal cortex was measured in 10-minute sessions from a total of $N = 7$ transgenic mice expressing the red-shifted calcium indicator jRGECO1a primarily in excitatory neurons.

We observed prominent, spontaneous fluctuations in pupil diameter and whisker motion throughout each recording session (Fig. 1B). Consistent with prior work [7, 72], these fluctuations were closely tracked by the leading linear dimension of the neural data, which we obtained via traditional principal component analysis (PCA) of the widefield calcium time series (Fig. 1B). Notably, this leading component dominated the calcium time series, accounting for between 70–90% of the variance in all seven mice (Fig. S1).

Such a dominant explanatory factor for global covariance has been widely interpreted to reflect "arousal"—a theoretical construct that, while rarely defined explicitly [34], is conventionally understood as a unidimensional continuum indexing the overall "wakefulness" or "activation" of an organism [24, 32, 73]. Although intuitively appealing, this framing tacitly promotes the view that brain activity naturally decomposes into linearly independent factors. In turn, this compels ad hoc explanations for the remaining orthogonal dimensions of neural data—dimensions that are, by definition, "not arousal". Here, these dimensions correspond to the remaining PCs (e.g., [7, 9, 10]); however, orthogonality also plays a central role in defining the scope of "arousal" in psychology via factor analysis (e.g., [74]), and in the common practice of linearly regressing out arousal indices.

An important limitation of these approaches is that linear dimensions need not correspond to physically meaningful variables. More fundamentally, a linear framework imposes a notion of independence that—while mathematically convenient—is essentially arbitrary from a biological perspective, and does not imply physical or mechanistic separability [75, 76] (see also [77, Ch. 6]).

Consistent with these points, we observed that higher PCs typically retained a clear temporal relationship to PC1, often closely approximating its successive time derivatives (Fig. 1B). These dynamical relationships are consistent with our alternative view: that scalar indices of "arousal", such as pupil size or a single PC, are best understood as one-dimensional projections of an underlying process—one whose full manifestations extend across multiple (Euclidean) dimensions (Fig. 1A). Crucially, on this view, the corresponding dimensions of the data remain fundamentally entangled, registering different aspects of a common process.

We refer to this underlying process as "arousal dynamics". Our core hypothesis is that the infra-slow spatiotemporal evolution of large-scale brain activity is primarily a reflection of this multidimensional, unified process. Below, we motivate this idea from a physiological perspective before introducing its dynamical systems framing. Ultimately, this framing enables us to empirically evaluate predictions motivated by this perspective.

Studying arousal dynamics

We interpret nominally arousal-related observables (e.g., pupil size, whisking) to reflect in large part the dynamics of a unified, organism-wide regulatory process (Fig. 1A). This dynamic regulation [78] may be viewed as a low-dimensional process whose evolution is tied to stereotyped variations in the high-dimensional state of the organism. Physiologically, this abstraction captures the familiar ability of arousal shifts to elicit multifaceted, coordinated changes throughout the organism (e.g., "fight-or-flight" vs. "rest-and-digest" modes) [79]. Operationally, this framing motivates a separation between the effective dynamics of this system, which can be represented on a latent, low-dimensional manifold [80–82], and the high-dimensional "observation" space, which we express via time-invariant mappings from the latent manifold. In this way, rather than assuming high-dimensional, pairwise interactions among the measured variables, we attempt to parsimoniously attribute widespread variance to a common, latent

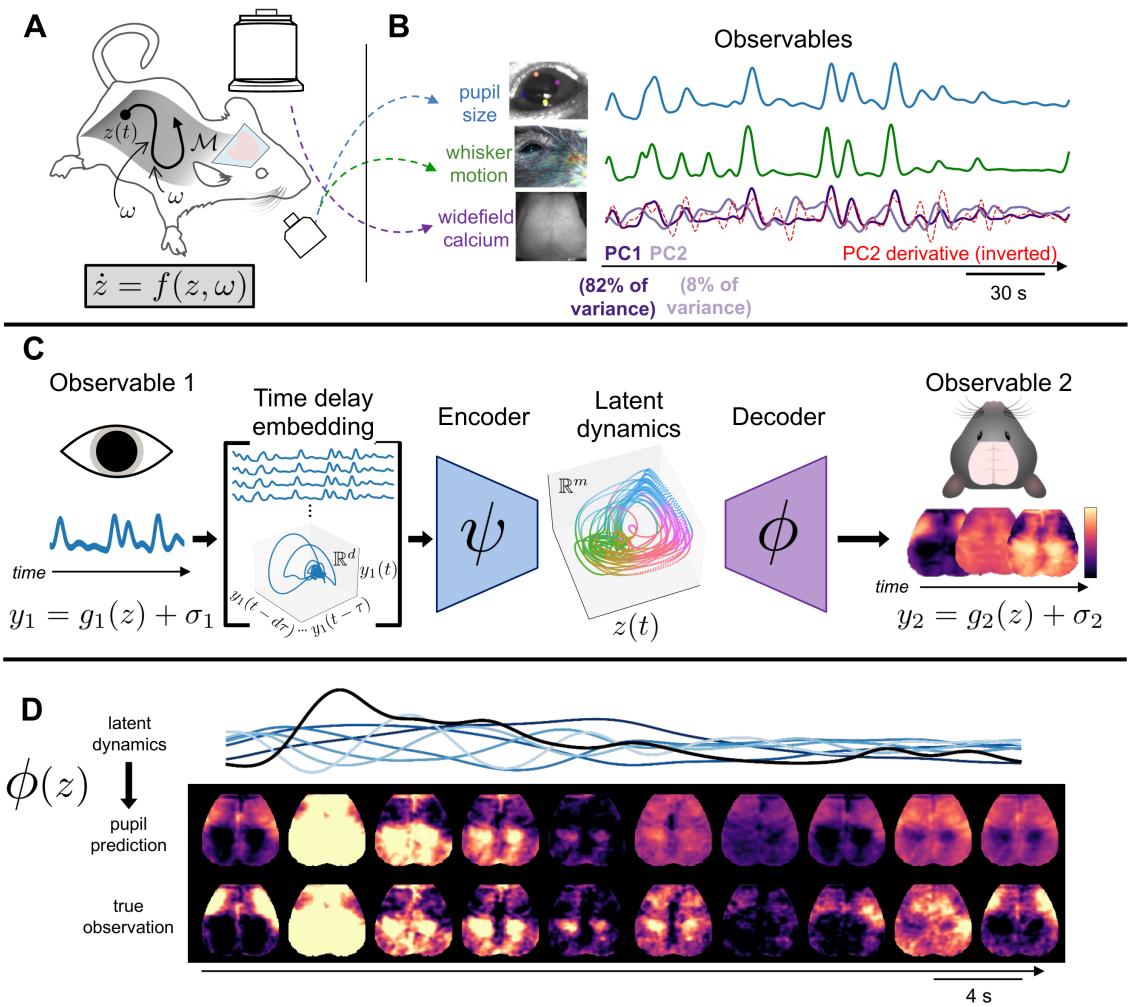

Figure 1: Study overview. **A** We performed simultaneous widefield optical imaging and face videography in awake mice. We interpret these measurements as observations of dynamics on an organism-intrinsic arousal manifold $\mathcal{M}$ (i.e., a low-dimensional surface within the state-space). Trajectories of the state z along $\mathcal{M}$ result from intrinsic arousal dynamics and various stochastic (i.e., arousal-independent) perturbations $\omega(t)$ (arising from within and outside the organism), such that $\dot{z} = f(z, \omega)$. **B** Three observable quantities derived from these recordings: scalar indices of pupil diameter and of whisker motion, and high-dimensional images of widefield calcium fluorescence. Here, widefield images are represented as the time series of the first two principal components (PCs). The first brain PC (PC1) closely tracked pupil diameter and whisker motion and accounted for 70–90% of the variance in each mouse (Fig. S1). However, remaining PCs often retained a clear temporal relationship with arousal. Specifically, here PC1 closely resembles PC2's derivative (red dashed line, Pearson $r = -.73$), implying that (nonlinear) correlation with arousal dynamics extends even beyond the dominant "arousal component" [7, 72]. **C** Our study describes an approach to parsimoniously model "arousal dynamics" and its manifestations. Each observable y_i is interpreted as the sum of an observation function on z and observation-specific variability σ_i (i.e., $y_i = g_i(z) + \sigma_i$). We test the specific hypothesis that pupil size and large-scale brain activity are both regulated by a common arousal-related process. Takens' embedding theorem [65] implies the possibility of relating these observables via a composition of encoder and decoder functions (i.e., ψ and ϕ) that map to and from a low-dimensional latent space representing the full state dynamics ($z(t)$). The encoder maps to this m-dimensional latent space from a $d-$dimensional "time delay embedding" ($d > m$) comprising d time-lagged copies of y_1 (i.e., pupil diameter). **D** An example epoch contrasting original and pupil-reconstructed widefield images (held-out test data from one mouse). Black is pupil diameter; blue traces are delay coordinates, where increasing lightness indicates increasing order (Methods).

dynamical system [83, 84].

To specify the instantaneous latent arousal state, our framework takes inspiration from Takens' embedding theorem [65], which provides general conditions under which the embedding of a scalar time series in time delay coordinates preserves the topological properties of (i.e., remains diffeomorphic to) the attractor in the full state-space of the original system. Delay embedding has become an increasingly powerful technique in modern data-driven dynamical systems [68]. Specifically, time delay coordinates have been shown to provide universal coordinates that represent complex dynamics for a wide array of observables [70], related to Koopman operator theory in dynamical systems [69, 85].

As we hypothesize that brain-wide physiology is spatiotemporally regulated in accordance with these arousal dynamics [63], Takens' theorem leads us to a surprising empirical prediction: a scalar observable of arousal dynamics (e.g., pupil diameter) can suffice to reconstruct the states of a high-dimensional

observable of brain physiology (e.g., optical images of cortex-wide activity), to the degree that the latter is in fact coupled to the same underlying dynamical process [67, 86, 87].

To test this hypothesis, we developed a framework for relating observables that evolve in time according to a shared dynamical system. Fig. 1C illustrates our data-driven approach to compute these relationships. In each mouse, we embed pupil diameter measurements in a high-dimensional space defined by its past values (i.e., time delay coordinates). We wish to discover a map (or encoder) ψ from this newly constructed observation space to a low-dimensional latent space z, which represents the instantaneous arousal state as a point within a continuous, observation-independent state-space. We additionally seek a mapping (decoder) ϕ from this latent space to the observation space of widefield calcium images. These mappings can be approximated with data-driven models learned using linear or nonlinear methods. Ultimately, this procedure enables us to express each instantaneous widefield calcium image as a function of a multidimensional state defined by pupil delay coordinates, to the extent that these two observables are deterministically linked to the same latent arousal state z.

Spatiotemporal dynamics from a scalar

For all analyses, we focused on ("infra-slow" [88, 89]) fluctuations <0.2 Hz, as this frequency range distinguishes arousal-related dynamics from physiologically distinct processes manifesting at higher frequencies (e.g., delta waves) [53, 90]. To model the hypothesized mappings, we implemented an iterative, leave-one-out strategy in which candidate models were trained using 10-minute recordings concatenated across six of the seven mice, and subsequently tested on the held-out 10-minute recording obtained from the seventh mouse. This modeling pipeline formed the basis for all main text results. We also implemented a separate, within-mouse pipeline in which we trained models on the first 5 minutes of data for each mouse and report model performance on the final 3.5 minutes of the 10-minute session. Results from this within-mouse pipeline are included in the Supplementary Appendix (Fig. S9). Details for both modeling pipelines are provided in the Methods.

For both pipelines, candidate models included a linear regression model based upon a single lag-optimized copy of pupil diameter ("No embedding"), as well as a model incorporating multiple delay coordinates and nonlinear mappings ("Latent model"). Importantly, both models allow us to assess the hypothesis that the widefield measurements primarily reflect arousal-related variation. However, the latter model adds the ability to capture complex aspects of pupil-widefield covariation that are more uniquely predicted by our dynamical view of arousal, enabling stronger validation of this particular interpretation. Nonlinear mappings were learned through simple feedforward neural networks, which were trained within an architecture modeled after the variational autoencoder [91] (with hyperparameters fixed across mice; Methods). See Supplementary Appendix (Supplementary Text and Figs. S2-S3) for validation of this framework on a toy (stochastic Lorenz) system.

Modeling results from the leave-one-out pipeline are shown in Fig. 2. We first considered the overall variance explained, summarized across time points and pixels. In each of $N = 7$ mice, we succeeded in reconstructing the majority of the variance in widefield calcium images on the basis of (simultaneously measured) pupil diameter, accounting for 60–85% of the total variance <0.2 Hz (Fig. 2A). As noted above, the total variance was dominated the first PC (Fig. S1), which showed a tight, linear relationship with pupil diameter. Accordingly, much of the total variance could also be modeled via linear regression with a one-dimensional pupil regressor (Fig. 2A). Thus, heuristically, both models support the notion that widefield calcium measurements are dominated by nominally "arousal-related" variation. However, as the total variance effectively treats each time point and pixel independently, this quantity fails to capture the dynamical and geometric structure inherent in the data—properties that may more directly relate to the underlying, spatiotemporally embedded generative process. Therefore, we next examined how well each model preserved these relational aspects of the data.

Crucially, we found that including multiple delays and nonlinearities additionally captured substantial variance orthogonal to the first PC (akin to residuals following global signal regression) (Fig. 2B) (see also Figs. S4, S5, and S9). Spatially, the topography of total explained variance was closely related to the first PC and strongly weighted toward posteromedial cortices (Fig. S6), resembling a familiar arousal-related topography as estimated by linear mappings from arousal indices [8, 9, 11, 92, 93]. However, in contrast, delays and nonlinearities allowed the model to also explain variance in anterolateral ("orofacial") cortices (Fig. S6), thus complementing the conventional arousal-related topography. The pattern of additional

explained variance resembled the spatial correlates of pupil derivatives and pupil amplitude envelope (Fig. S7)—two features which are implicitly represented in the delay embedding, and which are known to be tracked by neuromodulator dynamics [94, 95]. Interestingly, the largest increases in explained variance were consistently observed within a spatially focal set of regions that appeared to intersect the two main topographies (i.e., posteromedial and anterolateral cortices) (Fig. S6).

These consistent patterns suggest that modeling improvements did not trivially result from increased model complexity. Indeed, shuffling pupil and widefield measurements across mice precluded any ability to learn models that generalized to held-out data (Fig. S8). We conclude that an appreciable amount of widefield variance lying outside the dominant PC retains a time-invariant relationship with pupil dynamics (Fig. 2B-C).

The ability to model variance along multiple orthogonal dimensions of the widefield measurements indicates shared multidimensional, nonlinear structure between widefield and pupil dynamics. Nonetheless, this still does not directly address whether the model predictions faithfully reproduce the bona fide *spatiotemporal dynamics* of the widefield data, as would be expected if pupil and widefield measurements indeed share a common generative dynamical mechanism that is accurately captured by our model. To test this hypothesis more directly, we approximated propagation structure by fitting a linear dynamical systems model (via dynamic mode decomposition [69]; Methods) to both the original and reconstructed widefield measurements, then examined the leading spatial mode of each fitted dynamical system. In all seven mice, the original widefield data yielded a complex-valued leading spatial mode revealing a stereotyped pattern of cortical propagation consistent with prior observations [53, 96]. Crucially, data reconstructed by the "Latent model" consistently preserved this global propagation structure, as evidenced by the spatially similar phase patterns exhibited by the leading mode from each mouse (Fig. 2C). In contrast, the "No embedding" model failed to capture spatiotemporal dynamics altogether (see also Supp. Videos 1-2). This failure is expected, as bona fide spatiotemporal dynamics are inherently multidimensional: a one-dimensional dynamical system can yield only real-valued modes (eigenvectors), whereas true spatiotemporal dynamics correspond to interregional phase shifts that exist in the complex plane. Thus, generating and representing multiple distinct phase shifts necessitates at least two (Euclidean) dimensions.

Analyses based on hidden Markov models (HMMs) further supported the unique ability of delay embeddings to preserve the dynamical structure of the original widefield measurements (Fig. S10). Collectively, these findings and theoretical insights indicate that multiple orthogonal dimensions of brain measurements are inherently coupled due to an underlying spatiotemporal process—one that is faithfully reflected in the continuous evolution of the pupil.

Having successfully accounted for the majority of the variance, along with multidimensional and dynamical structure, we next asked whether these results could be attributed to high-amplitude, temporally isolated events, or whether they reflect a continuous coupling between the spatial topography of calcium measurements and the instantaneous state of arousal dynamics. To address this question, we evaluated the correspondence between the original and reconstructed datasets on a frame-to-frame basis, giving each time point (and thus, each image frame) equal weighting. Specifically, we first normalized each image frame to unit variance and projected it onto the top three PCs, then assigned these normalized, dimensionally-reduced images to clusters by fitting a Gaussian mixture model (GMM) (Methods). In this way, image frames were assigned to clusters based purely on their spatial similarity. Following this procedure, we assessed whether pupil dynamics allowed us to decode the cluster assignment of widefield data on a frame-to-frame basis.

Fig. 2D-E illustrates the results of these clustering and decoding analyses. Fig. 2D demonstrates the segregation of distinct spatial topographies to different regions of the arousal latent space (derived from pupil diameter), indicating that arousal dynamics are informative of spatially-specific patterns of cortex-wide activity. Fig. 2E shows decoding accuracy in test data as a function of the number of allowable clusters. Decoding accuracy was computed for each model as the percentage of "ground truth" GMM cluster assignments that matched the predictions generated by the same GMM conditioned on the reconstructed (rather than the original) image frames. A "null" score was obtained by assigning all frames to the most frequent assignment in the ground truth data.

In general, we observed that spatial patterns largely alternated between two primary patterns, consistent with large-scale spatiotemporal measurements reported across modalities and species [62, 63, 92, 97]. With a two-cluster solution, pupil dynamics enabled accurate assignment of more than 80% of

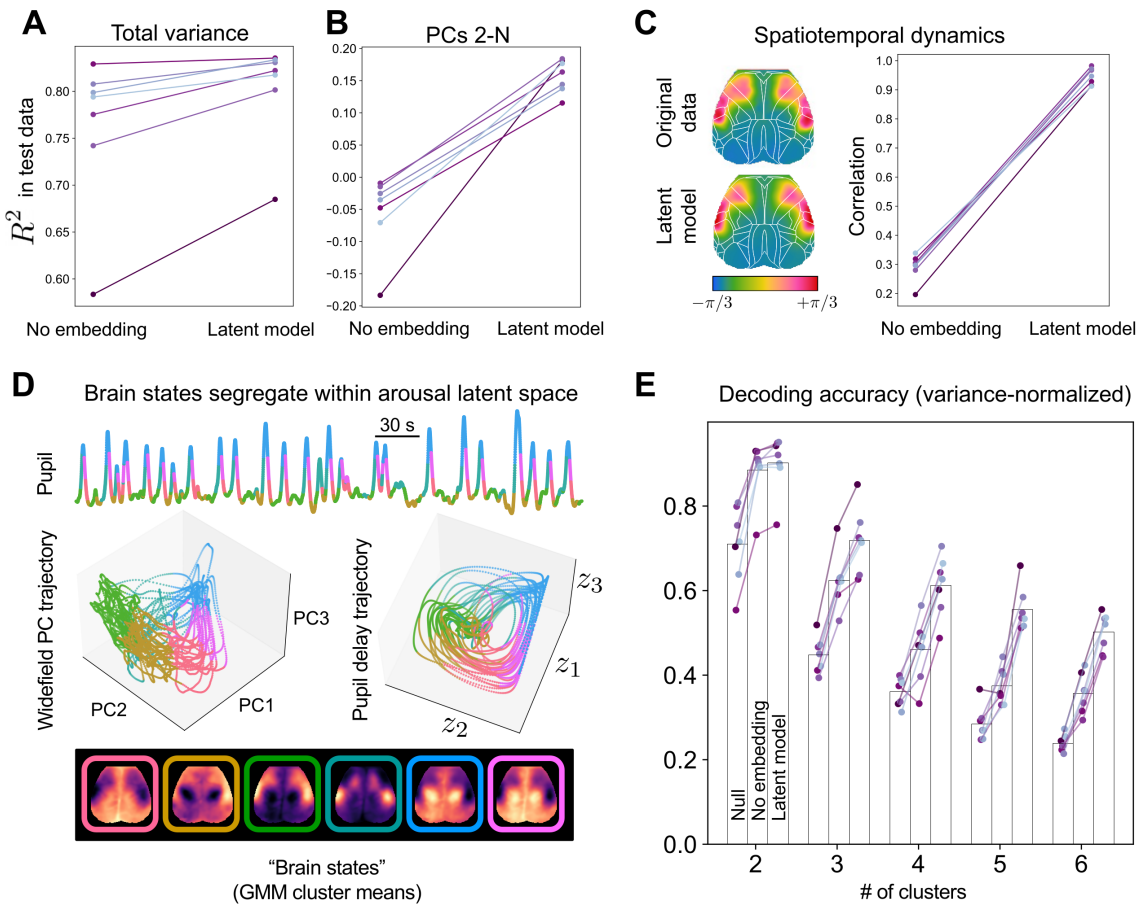

Figure 2: Pupil diameter enables reconstruction of spatiotemporal brain dynamics. A Variance explained in held-out widefield data from each mouse (for each prediction, the model was trained on data from the other 6 mice). A simple linear regression model based on one (lag-optimized) pupil regressor ("No embedding") accurately predicted $\sim$ 70–85% of variance in six of seven mice. Delay embedding and nonlinearities ("Latent model") extended this level of performance to all mice while further augmenting model performance, primarily by accounting for variance spread along higher dimensions (PCs 2-N) (B). See Fig. S6 for pixelwise maps of variance explained. **C** Analysis of propagation dynamics. Phase maps were obtained from dynamic mode decomposition applied to the original and reconstructed data, and subsequently compared via (circular) spatial correlation. Phase maps from one example mouse are shown on the left (phase maps from the 1D model are not shown, as they lack phase structure due to the absence of complex-valued modes). Allen Mouse Common Coordinate Framework atlas overlaid in white. **D** Example state trajectories (held-out data) visualized in coordinates defined by pupil diameter (top), the first three widefield PCs (left) or three latent variables obtained from the delay-embedded pupil time series (right). Trajectories are color-coded by cluster identity as determined by a Gaussian mixture model (GMM) applied to widefield image frames ($k = 6$ clusters). Cluster means are shown below these trajectory plots. Color segregation in pupil delay coordinates indicates spatially-specific information encoded in pupil dynamics. Similar color arrangement within both embeddings suggests their shared topological properties. **E** Decoding "brain states" from pupil dynamics. Comparison of decoder accuracy in properly identifying GMM cluster assignments for each image frame based on the most frequent assignment in the training set ("Null" decoder), the "No embedding" model, or the "Latent model".

image frames (Fig. 2E). Decoding accuracy remained high with three distinct topographies, with the latent model enabling an average accuracy of >70%. With more than three clusters, we found spatial topographies to be increasingly less distinctive, leading to gradually worsening decoding accuracy. Yet, decoding accuracy continued to follow a common trend, steadily improving from the "Null" decoder to the "No embedding" decoder, and, finally, to the "Latent model" decoder. The superior accuracy of the latter was primarily rooted in its ability to simultaneously track spatial variation along multiple orthogonal dimensions, rather than just the first PC. Indeed, discrepancies in prediction accuracy became even more pronounced when assigning equal weight to the three PCs used for the clustering analysis (Fig. S11). In addition, we observed similar or stronger decoding advantages of the latent model when labeling states according to an HMM rather than a GMM (Fig. S10). These results confirm that pupil dynamics reliably index the instantaneous spatial topography of large-scale calcium fluorescence.

Collectively, our analyses support the interpretation that arousal dynamics, as indexed by pupil diameter, represent the predominant physiological process underlying cortex-wide spatiotemporal dynamics on the timescale of seconds.

Multimodal spatiotemporal dynamics from a shared latent space

We have hypothesized that arousal dynamics support the spatiotemporal regulation of brain-wide *physiology*—thus extending beyond the activity of neurons per se (e.g., [29, 72, 98]). Spatially structured fluctuations in metabolic and hemodynamic measurements are also well-established, and are generally interpreted to reflect state-, region-, and context-dependent responses to local neuronal signaling [99]. On the other hand, our account posits that multiple aspects of physiology are jointly regulated according to a common process. This suggests the possibility of more parsimoniously relating multiple spatiotemporal readouts of brain physiology through a common low-dimensional latent space reconstructed from pupil diameter.

To test this prediction, we took advantage of the additional readouts provided by the multispectral optical imaging platform [71]—namely, oxidative metabolism (as indexed by flavoprotein autofluorescence, or FAF [100–102]) and hemodynamics (which approximates the physiological basis for fMRI), which were obtained simultaneously with the neuronal calcium measurements. We attempted to reconstruct all three observables on the basis of delay embedded pupil measurements. Importantly, we did not retrain the encoder ψ learned in the context of predicting calcium measurements in Fig. 2. Rather, we simply trained additional decoders to map from the latent space defined by ψ to spatiotemporal measurements of FAF and oxygenated hemoglobin (ϕ_2 and ϕ_3 in Fig. 3A, respectively).

Modeling accuracy for FAF and hemoglobin measurements are shown in Fig. 3B. As with calcium fluorescence, we were able to accurately model the large-scale spatiotemporal dynamics of these two measurements in each of $N = 7$ mice. These measurements were similarly dominated by the first PC (Figs. S12 & S13). Interestingly, however, for these measurements, delay embeddings and nonlinearities made a much larger contribution to the explained variance, even after adjusting for a globally-optimized delay (Figs. 3C & S14). Thus, a far larger portion of metabolic and hemodynamic fluctuations can be attributed to a common, arousal-related mechanism than is conventionally recognized (see also [54, 63]). For context, predictions based on pixelwise calcium signals are shown in Fig. S15). We revisit the physiological interpretation of these relationships in the Discussion.

Unifying observables through arousal dynamics

Analyses thus far support our dynamical systems interpretation of arousal, which offers a parsimonious account of ongoing fluctuations in multimodal measurements of large-scale spatiotemporal brain activity. A key motivation for this framing is the growing diversity of observables linked to arousal indices [6], despite the persistent lack of clarity surrounding this construct [34]. Accordingly, we now consider how our framework can quantitatively integrate such findings via a unified representation of arousal.

Crucially, by relating observables through an organism-intrinsic manifold, rather than an imposed stimulus or task (cf. [103]), our framework has the potential to generalize across different experimental datasets, enabling integration of our findings into a more holistic context than is feasible in any single experimental study. To explore this potential, we extended our framework to publicly available behavioral and electrophysiological recordings from the Allen Institute Brain Observatory [104]. This dataset includes 30-minute recording sessions of mice in a task-free context, i.e., without explicit experimental intervention. From this dataset, we computed eight commonly used indices of brain state and arousal (see Methods): pupil diameter, running speed, hippocampal θ/δ ratio, the instantaneous rate of hippocampal sharp-wave ripples (SWRs), bandlimited power (BLP) derived from the local field potential (LFP) across visual cortical regions, and mean firing rate across several hundred neurons per mouse recorded with Neuropixels probes [105]. The two hippocampal indices were both derived from the LFP of the CA1 region. BLP across visual cortex was further analyzed within three canonical frequency ranges: 0.5–4 Hz ("delta"), 3–6 Hz ("alpha" [106]), and 40–100 Hz ("gamma").

We began by concatenating measurements across mice, then trained neural networks to predict each observable from a latent space constructed from (delay embedded) pupil measurements. After training on concatenated measurements, we applied the learned encoding to the delay embedded pupil measurements of each individual mouse, training decoders to map from this shared latent space to the mouse's own observables. Finally, we repeated this procedure for the primary widefield dataset—thus using pupil diameter as a common reference to align observables across datasets (and mice) to a shared manifold (see Methods for further details).

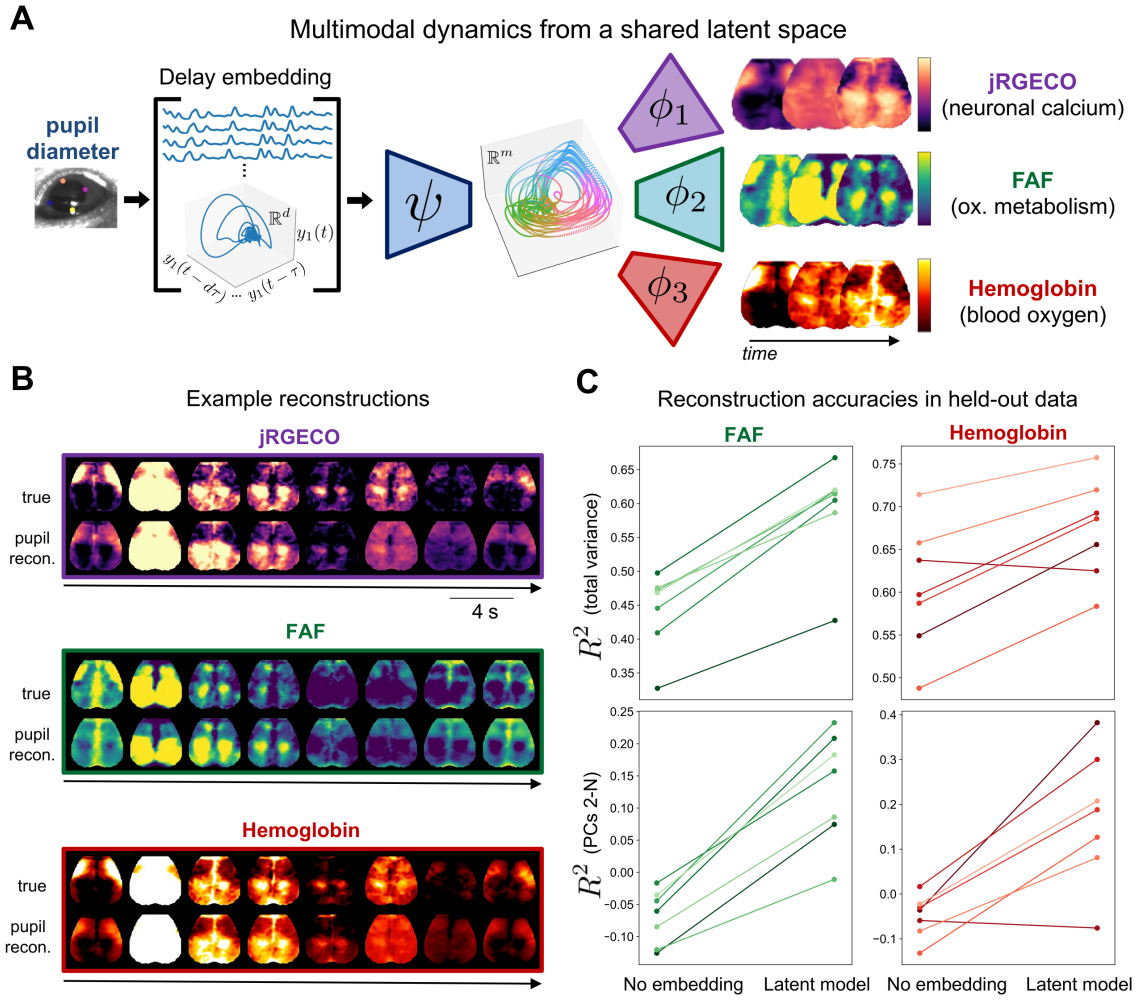

Figure 3: Multimodal physiological measurements can be reconstructed from a shared latent manifold. A. The delay embedding framework extended to multimodal observables. We used the same encoder previously trained to predict calcium images. We coupled this encoder to decoders trained separately to predict FAF and hemoglobin. **B.** Example reconstructions of neuronal calcium (jRGECO fluorescence), metabolism (FAF), and blood oxygen (concentration of oxygenated hemoglobin). **C.** Reconstruction performance for each observable (as in Fig. 2A). See Fig. 2 for results with calcium, and Fig. S14 for pixelwise maps of variance explained.

Fig. 4 illustrates the resulting, multi-dataset generative model. For visualization, we defined the arousal manifold $\mathcal{M}$ as a 2D latent space parameterized by two discovered coordinates, z_1 and z_2. Next, we characterized $\mathcal{M}$ in terms of each observable y_i by evaluating the corresponding mappings ϕ_i at each point within a 2D grid of sample points within the latent space. Finally, we obtained a data-driven analytic representation of the latent dynamics [107] (Methods), allowing us to compute the vector field $f(z)$ at each point within the 2D grid. Taken together, at each point along $\mathcal{M}$, Fig. 4 depicts the conditional expectation for (1) the direction of flow of the latent arousal dynamics (via f), and (2) the state of each observable y_i (via ϕ_i). In this way, $\mathcal{M}$, f, and the mappings $\{\phi_i\}_{i=1}^{N}$ together encapsulate our dynamical, relational interpretation of arousal.

The resulting representation unifies and clarifies relations among several themes appearing in different sectors of the experimental literature, which we briefly summarize here. As expected, running speed, hippocampal θ/δ ratio, cortical gamma BLP, and mean firing rate are all positively correlated with pupil diameter along a single, principal dimension, z_1 (i.e., the horizontal coordinate in each plot) (Fig. 4, "Allen Brain Observables"). In contrast, delta BLP, alpha BLP, and hippocampal SWR rate—all conventionally associated with internally-oriented functionalities and momentary sensory disengagement—are inversely correlated with pupil diameter along the same z_1 dimension [13, 108, 109].

A second dimension (z_2, vertical coordinate in each plot) clarifies more subtle temporal relationships, revealing a canonical cycle that captures diverse phenomenology previously reported in the context of brain and behavioral state transitions [94, 95, 110, 111]. Thus, the "arousal onset" (Fig. 4) begins with

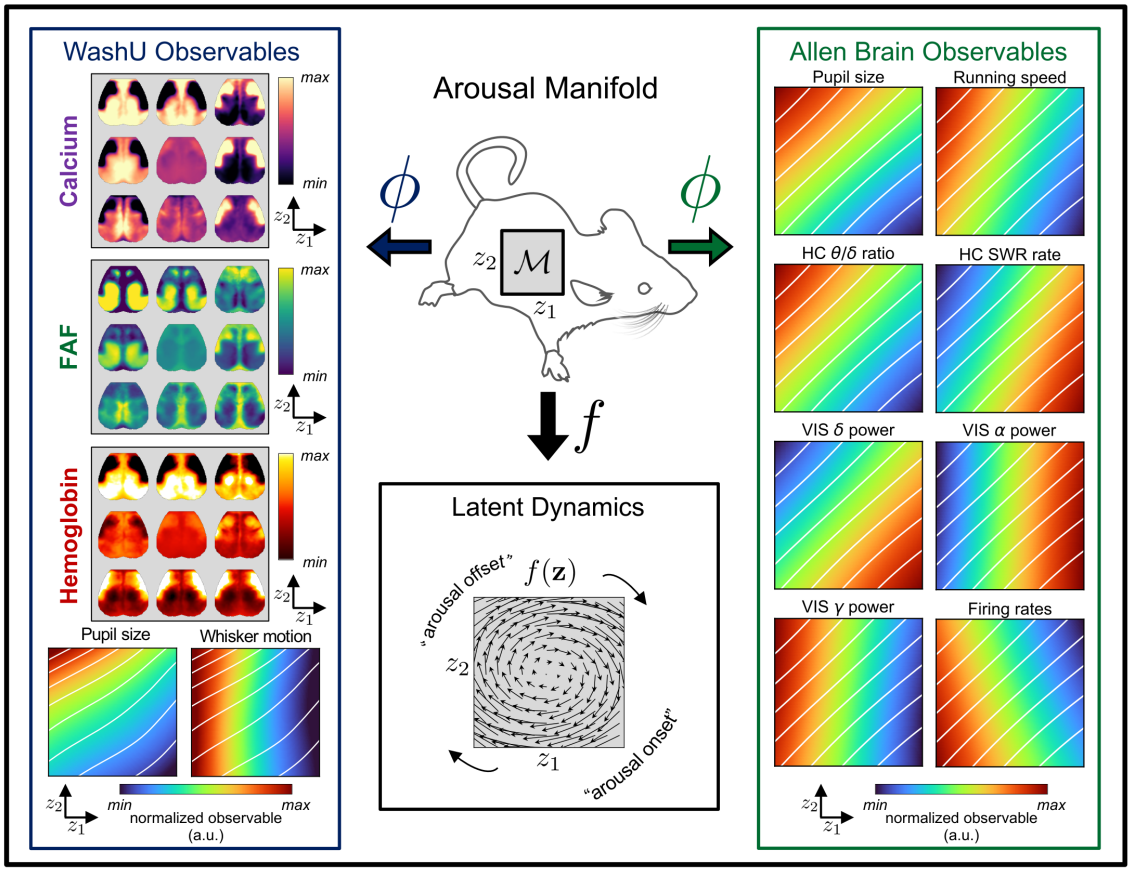

Figure 4: Unifying observables through arousal dynamics. The arousal manifold provides an organism-intrinsic reference frame to align and aggregate measurements across datasets. To visualize the discovered relations, we begin by defining the arousal manifold as a 2D latent space parameterized by the coordinates z_1 and z_2 (i.e., $\mathbf{z} \in \mathcal{M} \subset \mathbb{R}^2$) (upper middle). Each point $\mathbf{z}$ in this 2D manifold represents a latent arousal state. The learned decoders ϕ map each of these points to an expected value of a particular observable (e.g., a single widefield calcium image, or a single value of pupil size), thus enabling representations of the 2D arousal manifold from the perspective of various observables (left and right columns). For visualization, the widefield observables are evaluated at 3×3 evenly spaced points within the 2D latent space, whereas each scalar observable is evaluated over a grid of 100×100 sample points). White diagonals indicate iso-contours of pupil size in either the WashU or Allen Brain Observatory datasets, allowing a common visual reference across observables within the same dataset. Meanwhile, $f(\mathbf{z})$, which we approximate via a data-driven analytic representation [107], maps each point in the latent manifold to an expected direction of movement (i.e., $\dot{\mathbf{z}} = f(\mathbf{z})$), enabling the manifold to be represented also as a vector field (lower middle). Together, these representations depict how a typical sequence of arousal states, as determined by the latent, intrinsic arousal dynamics, would manifest in terms of changes to observable quantities. For instance, beginning at "arousal onset" (lower right corner of each square in the figure), the clockwise rotation indicated by arrows in $f(\mathbf{z})$ is expected to manifest as an increase and subsequent decrease in whisker motion; similar changes are expected in pupil size, but slightly delayed, as the max pupil size is rotated clockwise from max whisker motion. FAF, flavoprotein autofluorescence; HC, hippocampus; SWR, sharp-wave ripple; VIS, visual cortex.

decreasing alpha BLP [111] and subsequent increases in firing rates and gamma BLP, which precede an increase in pupil diameter [72, 110]. The "WashU observables" indicate that this arousal onset period is also characterized by the onset of whisking and a propagation of cortical activity toward posterior regions [19, 53, 92], with calcium changes briefly lagged by metabolic and then hemodynamic changes.

Continuing clockwise from this "arousal onset" period, the Allen Brain Observables indicate roughly synchronous increases in pupil diameter, locomotion, and hippocampal θ/δ ratio. After these observables reach their maximum, alpha BLP begins to increase, consistent with the appearance of this cortical oscillation at the offset of locomotion [106] (marked as "arousal offset" in Fig. 4). This is succeeded by a general increase in delta BLP and hippocampal SWR rate. Across the cortex, this behavioral and electrophysiological sequence at "arousal offset" (as indicated by the Allen Brain Observables) is associated with a propagation of activity from posterior to anterior cortical regions (as indicated by the WashU observables). Finally, activity appears across lateral orofacial cortices, which remain preferentially active until the next arousal onset. This embedding of cortex-wide topographic maps within a canonical "arousal cycle" resembles prior results obtained from multimodal measurements in humans and monkeys, which additionally link these cortical dynamics to topographically parallel rotating waves in subcortical structures [63].

Taken together, this framework provides a unified, parsimonious, and quantitative representation of a functionally coherent spatiotemporal pattern of changes that is registered only piecemeal through diverse experimental modalities. Although verifying this model would ideally require simultaneous measurements of all observables—which remains practically infeasible—the ability to integrate partial observations from different experiments plays a key role in contextualizing disparate observations, facilitating hypothesis generation, and enabling large-scale model construction, thereby laying groundwork for future experiment-theory iteration.

Discussion

We have proposed to interpret arousal as a specific, multidimensional process that spatiotemporally regulates brain-wide physiology in parallel with organismal physiology and behavior. We support this account through a data-driven modeling framework, demonstrating that a scalar index of arousal, pupil diameter, suffices to reconstruct the multidimensional state dynamics underlying the spatiotemporal evolution of cortex-wide physiology. We interpret these results as support for the hypothesis that spontaneous, infraslow fluctuations in large-scale brain activity primarily reflect an intrinsic regulatory mechanism [63]. Collectively, these insights carry important implications for understanding brain states and arousal, large-scale brain dynamics, and the interrelations among multimodal readouts of brain physiology. These implications are considered in what follows.

The concept of "brain state" has gained heightened significance in systems neuroscience as experimental advances have made studies in awake, behaving organisms increasingly routine [5, 6, 35]. Although brain state is typically viewed as comprising several biological, cognitive, or behavioral factors—one of which is often presumed to be "arousal" [1, 10, 35]—it has remained unclear how to reliably decompose brain state without more precisely defined factors. Construing organisms as dynamical systems [83, 112], our approach has been to factorize brain (or rather, organismal) state in terms of constituent endogenous processes—one of which, examined herein, we identify with the arousal construct. Despite grounding arousal in a specific process, our results reveal this process to be multidimensional and nonlinearly coupled to various observables—ultimately implying even broader relevance for arousal than presently recognized.

Our generalized view of arousal resonates with two emerging themes in the literature. First, brain or arousal states appear to represent distinct (neuro-)physiological regimes, such that state changes are associated with sweeping changes in the behavior of faster timescale processes [29, 113, 114]. This limits the ability to account for arousal-dependence by simply orthogonalizing observations with respect to a scalar arousal index (e.g., via linear regression of pupil diameter or removal of the leading PC; Fig. 2B). Second, numerous studies have demonstrated non-random transitions among brain states [1, 12, 115, 116], along with the functional relevance of *dynamical* indices of arousal, such as the phase or derivative of pupil size [72, 94, 110, 117], often with surprising timescale-invariance [14, 63, 90]. These observations imply an underlying, nonlinear dynamic process [118]. Together, these two themes suggest a view of arousal as constituting a complex, pervasive, and persistent dynamic "background"; this complicates the ability to interpret spontaneous or task-related [9, 11] variation observed over similar timescales—that is, even on the order of seconds—without reference to this changing *internal context* [2, 17]. To this end, modeling approaches such as in Fig. 4 can refine interpretations by more rigorously conditioning observations upon this nonlinear dynamic background.

Indeed, by more rigorously capturing the scope of arousal dynamics, our models reveal surprisingly rich spatiotemporal structure across the cortex, prompting reconsideration of the mechanisms typically assumed responsible. Thus, over decades of research spanning multiple experimental communities, large-scale spatiotemporal brain dynamics have been overwhelmingly viewed as summations of diverse interregional interactions [37, 39, 41, 119], fueling a proliferation of decomposition techniques. Although distinctions among modalities and species should not be overlooked, evidence across these domains has increasingly converged toward a sparse set of core underlying spatiotemporal components [52, 56, 62], alongside growing appreciation for arousal-related contributions [31, 43, 46, 51]. We extend these themes by advocating a more radically unified view: that these core spatiotemporal components are themselves projections of a nondecomposable dynamical system, and that this system effectively underpins arousal. In this sense, arousal dynamics directly account for the observed spatiotemporal structure—i.e., not (primarily) by reconfiguring interregional neuronal interactions, but by actively generating spatiotempo-

ral structure through the synchronized [87] propagations of multiple topographically organized rotating waves [63].

Importantly, our unified model specifically concerns infra-slow spatiotemporal brain dynamics. Significantly faster processes (particularly those above 1 Hz) are effectively decoupled from this timescale and likely involve distinct physiological mechanisms [90], including the kinds of flexible and regionally specific interactions that we argue play a smaller role at infra-slow timescales. Interestingly, widefield optical recordings reveal similar spatial patterns across both timescales [120], likely reflecting shared anatomical constraints (in our model, such constraints are implicitly subsumed by the time-invariant mapping ϕ). Nonetheless, temporal dynamics among these spatial patterns seem to differentiate the two timescales [53, 96], supporting their distinct underlying mechanisms. Still, rather than being entirely separate, spatiotemporal dynamics above 1 Hz are strongly modulated by infra-slow (and/or arousal) dynamics [53, 120, 121]. Likewise, a rich literature describes the amplitude modulation of higher-frequency brain activity according to the phase of infra-slow fluctuations [89]. We suggest that these multiscale relationships are special cases of a more general pattern by which the infra-slow evolution of arousal state modulates the parameters of faster timescale processes throughout the brain and body. Indeed, the conventional brain state indices in Fig. 4 can be viewed as precisely such parameters (e.g., "rate" and "power"). This pattern suggests multiscale extensions of the present framework, wherein dynamical models describing faster processes can be jointly parameterized by the state of a common arousal process.

This theoretical framing also leads us to a simplified view of the interrelations between electrical, metabolic, and vascular cerebral activity. Although brain function necessitates tight coupling among all three aspects of physiology, their intersection constitutes an extraordinarily high-dimensional space of biochemical interactions. These interactions are commonly viewed in terms of complex, region-dependent, and still incompletely understood neurometabolic and neurovascular cascades operating locally throughout the brain [99], posing an enormous regulatory challenge. On the other hand, there is longstanding evidence that neuromodulatory processes can directly influence neural, metabolic, and vascular activities in parallel [122–125], suggesting mechanisms enabling their coordinated regulation. Based on the strong performance of our model (Fig. 3), we propose that such mechanisms are actively harnessed to effectively streamline cerebral regulation in accordance with arousal dynamics—for instance, through spatiotemporally coordinated changes to high-level parameters such as neuronal gain [126] and metabolic fluxes [127]. This strategy would reduce reliance on more conventionally studied feedforward responses to local neuronal activity (e.g., "the hemodynamic response"), which may come secondary to more integrative mechanisms supporting the continuous, dynamic, and globally coordinated nature of brain physiology (see also [29, 128]). This interpretation could help explain observations that appear more puzzling when viewing spontaneous hemodynamic fluctuations as mere responses to local neuronal activity (e.g., [129]). Ultimately, targeted experiments are needed to determine the extent to which neuromodulatory processes directly and indirectly contribute to neurometabolic and neurovascular coupling. Regardless, our results indicate that the observed dynamics and relationships are consistent with widespread coupling to a global regulatory process.

Finally, although we have adopted the term "arousal" given widespread adoption of this term for the observables discussed herein, our results are obtained through an entirely data-driven approach: "arousal" is simply the label we assign to the latent process inferred from the observables. In other words, we make no a priori assumptions about how arousal might be represented in these observables in comparison to other behavioral processes or states [130]. For this same reason, our results are likely relevant to a range of cognitive and behavioral processes that are traditionally distinguished from arousal (e.g., "attention"), but which may be intimately coupled (or largely redundant) with the dynamical process formalized herein [2, 131].

Methods

Datasets and preprocessing

Dataset 1: Widefield optical imaging (WashU)

Animal preparation All procedures described below were approved by the Washington University Animal Studies Committee in compliance with the American Association for Accreditation of Laboratory Animal Care guidelines. Mice were raised in standard cages in a double-barrier mouse facility with a 12 h–12 h light/dark cycle and ad libitum access to food and water. Experiments used N=7 12-week old mice hemizygous for Thy1-jRGECO1a (JAX 030525) on a C57BL/6J background, enabling optical imaging of the jRGECO1a fluorescent calcium sensor protein primarily expressed in excitatory neurons of cortical layers 2/3 and 5 [132]. Prior to imaging, a cranial window was secured to the intact skull of each mouse with dental cement under isoflurane anesthesia according to previously published protocols [133].

Data acquisition Widefield imaging was conducted on a dual fluorophore optical imaging system; details of this system have been presented elsewhere [71, 134]. Briefly, sequential illumination was provided at 470 nm, 530 nm, and 625 nm; reflected light in each channel was collected by a lens (focal length = 75mm, NMV-75M1, Navitar), split by a 580 nm dichroic (FF580-FDi02-t3-25x36, Semrock) into two channels, one filtered by 500 nm long pass (FF01-500/LP-25, Semrock – blocks FAF excitation, passes FAF emission and 530 nm Hb reflectance), the other by a 593 nm long pass (FF01-593/LP-25, Semrock - blocks jRGECO1a excitation, passes jRGECO1a emission and 625 nm Hb reflectance). The two channels were detected separately via two CMOS cameras (Zyla 5.5, Andor). Data were cropped to 1024x1024 pixels, and binned to 512x512 to achieve a frame rate of 100Hz in each camera, with all contrasts imaged at 25 Hz. The LEDs and camera exposures were synchronized and triggered via a data acquisition card (PCI-6733, National Instruments) using MATLAB R2019a (MathWorks).

Mice were head-fixed under the imaging system objective using an aluminum bracket attached to a skull-mounted Plexiglas window. Prior to data acquisition in the awake state, mice were acclimated to head-fixation over several days.

The mouse's body was supported by a felt pouch suspended by optical posts (Thorlabs). Resting state imaging was performed for 10 minutes in each mouse. Before each imaging run, dark counts were imaged for each mouse for 1 second with all LEDs off in order to remove background sensor noise.

Preprocessing Images were spatially normalized, downsampled to 128x128 pixels, co-registered, and affine-transformed to the Paxinos atlas, temporally detrended (5th order polynomial fit) and spatially smoothed (5x5 pixel Gaussian filter with standard deviation of 1.3 pixels) [135]. Changes in 530 nm, and 625 nm reflectance were interpreted using the modified Beer-Lambert law to calculate changes in oxy- and deoxy-hemoglobin concentration [71].

Image sequences of fluorescence emission detected by CMOS1 (i.e., uncorrected FAF) and CMOS2 (i.e., uncorrected jRGECO1a) were converted to percent change (dF/F) by dividing each pixel's time trace by its average fluorescence over each imaging run. Absorption of excitation and emission light for each fluorophore due to hemoglobin was corrected as outlined in [136]. From the face videography we derived scalar indices of pupil size via DeepLabCut software [137] and whisker motion via the Lucas-Kanade optical flow method [138] applied to and subsequently averaged across five manually selected data points on the whiskers. The resulting pupil diameter, whisker motion, and widefield time series were bandpass filtered between $0.01 < f < 0.2$ Hz in order to distinguish the hypothesized spatiotemporal process from distinct phenomena occurring at higher frequencies (e.g., slow waves) [53], and to accommodate finite scan duration 10 minutes. For visualization only—namely, to view the Allen atlas boundaries in register with cortical maps—we manually aligned the Paxinos-registered data from one mouse to a 2D projection of the Allen Mouse Common Coordinate Framework v3 [139] using four anatomical landmarks: the left, right, and midline points at which the anterior cortex meets the olfactory bulbs, and an additional midline point at the base of retrosplenial cortex [9, 140]. Coordinates obtained from this transformation were used to overlay the CCF boundaries onto the Paxinos-registered cortical maps for all mice.

Dataset 2: Allen Brain Observatory

We additionally analyzed recordings obtained from awake mice and publicly released via the Allen Brain Observatory Neuropixels Visual Coding project [104]. These recordings include eye-tracking, running speed (estimated from running wheel velocity), and high-dimensional electrophysiological recordings obtained with Neuropixels probes [105]. We restricted analyses to the "Functional Connectivity" stimulus set, which included a 30-minute "spontaneous" session with no overt experimental stimulation. Data were accessed via the Allen Software Development Kit (SDK) (https://allensdk.readthedocs.io/en/latest/). From 26 available recording sessions, we excluded four with missing or compromised eye-tracking data, one lacking local field potential (LFP) data, and five with locomotion data indicating a lack of immobile periods or appearing otherwise anomalous, thus leaving a total of 16 recordings for downstream analyses.

Preprocessing From the electrophysiological data we derived estimates of population firing rates and several quantities based upon the LFP. We accessed spiking activity as already extracted by Kilosort2 [7]. Spikes were binned in 1.2 s bins as previously [7], and a mean firing rate was computed across all available units from neocortical regions surpassing the default quality control criteria. Neocortical LFPs were used to compute band-limited power (BLP) within three canonical frequency bands: gamma ($40 - 100$ Hz), "alpha" ($3 - 6$ Hz) and delta ($0.5 - 4$ Hz). Specifically, we downsampled and low-passed the signals with a 100 Hz cutoff, computed the spectrogram (*scipy.signal.spectrogram* [141]) at each channel in sliding windows of 0.5 s with 80% overlap, and averaged spectrograms across all channels falling within "VIS" regions (including primary and secondary visual cortical areas). BLP was then computed by integrating the spectrogram within the frequency bands of interest and normalizing by the total power. A similar procedure was applied to LFP recordings from the hippocampal CA1 region to derive estimates of hippocampal theta ($5 - 8$ Hz) and delta ($0.5 - 4$ Hz) BLP. Hippocampal sharp-wave ripples were detected on the basis of the hippocampal CA1 LFP via an automated algorithm [142] following previously described procedures [143]. In general, extreme outlying values (generally artifact) were largely mitigated through median filters applied to all observables, whereas an additional thresholding procedure was used to interpolate over large negative transients in running speed. Finally, all time series were downsampled to a sampling rate of 20 Hz and filtered between $0.01 < f < 0.2$ Hz to facilitate integration with Dataset 1.

Data analysis

State-space framework

Formally, we consider the generic state-space model:

$$\dot{\mathbf{z}} = \mathbf{f}(\mathbf{z}, \omega), \tag{1}$$

$$\mathbf{y}_i = \mathbf{g}_i(\mathbf{z}) + \sigma_i, \tag{2}$$

where the nonlinear function $\mathbf{f}$ determines the (nonautonomous) flow of the latent (i.e., unobserved), vector-valued arousal state $\mathbf{z} = [z_1, z_2, \ldots, z_m]^{\mathrm{T}}$ along a low-dimensional attractor manifold $\mathcal{M}$, while $\omega(t)$ reflects random (external or internal) perturbations decoupled from $\mathbf{f}$ that nonetheless influence the evolution of $\mathbf{z}$ (i.e., "dynamical noise"). We consider our N observables $\{\mathbf{y}_i\}_{i=1}^{N}$ as measurements of the arousal dynamics, each resulting from an observation equation $\mathbf{g}_i$, along with other contributions $\sigma_i(t)$ that are decoupled from the dynamics of $\mathbf{z}$ (and in general are unique to each observable). Thus, in this framing, samples of the observable $\mathbf{y}_i$ at consecutive time points t and $t+1$ are linked only through the evolution of the latent variable $\mathbf{z}$ as determined by $\mathbf{f}$. Note that $\mathbf{f}$ is presumed to embody various causal influences spanning both brain and body (and the feedback between them), whereas *measurements* (of either brain or body) $\mathbf{y}_i$ do not contribute to these dynamics. This framework provides a formal, data-driven approach to parsimoniously capture the diverse manifestations of arousal dynamics, represented by the proportion of variance in each observable $\mathbf{y}_i$ that can be modeled purely as a time-invariant mapping from the state-space of $\mathbf{z}$.

State-space reconstruction: time delay embedding

Our principal task is to learn a mapping from arousal-related observables to the multidimensional space where the hypothesized arousal process evolves in time. To do this, we take advantage of Takens' embedding theorem from dynamical systems theory, which has been widely used for the purpose of nonlinear state-space reconstruction from an observable. Given p snapshots of the scalar observable y in time, we begin by constructing the following Hankel matrix $\mathbf{H} \in \mathbb{R}^{d\times(p-d+1)}$:

$$\mathbf{H} = \begin{bmatrix} y(t_1) & y(t_2) & \dots & y(t_{p-d+1}) \\ y(t_2) & y(t_3) & \dots & y(t_{p-d+2}) \\ \vdots & \vdots & \ddots & \vdots \\ y(t_d) & y(t_{d+1}) & \dots & y(t_p) \end{bmatrix} \tag{3}$$

for p time points and d time delays. Each column of $\mathbf{H}$ represents a short trajectory of the scalar observable $y(t)$ over d time points, which we refer to as the *delay vector* $\mathbf{h}(t)$. These delay vectors represent the evolution of the observable within an augmented, $d-$dimensional state-space.

We initially construct $\mathbf{H}$ as a high-dimensional (and rank-deficient) matrix to ensure it covers a sufficiently large span to embed the manifold. We subsequently reduce the dimensionality of this matrix to improve conditioning and reduce noise. Dimensionality reduction is carried out in two steps—first, through projection onto an orthogonal set of basis vectors (below), and subsequently through nonlinear dimensionality reduction via a neural network (detailed in the next section).

For the initial projection, we note that the leading left eigenvectors of $\mathbf{H}$ converge to Legendre polynomials in the limit of short delay windows [144]. Accordingly, we use the first $r = 10$ discrete Legendre polynomials as the basis vectors of an orthogonal projection matrix:

$$\mathbf{P}^{(r)} = [\mathbf{p}_0, \mathbf{p}_1, \dots, \mathbf{p}_{r-1}] \in \mathbb{R}^{d\times r} \tag{4}$$

(polynomials obtained from the *special.legendre* function in SciPy [141]). We apply this projection to the Hankel matrix constructed from a pupil diameter timecourse. The resulting matrix $\tilde{\mathbf{Y}} \in \mathbb{R}^{\mathrm{r}\times(\mathrm{p}-\mathrm{d}+1)}$ is given by:

$$\tilde{\mathbf{Y}} = \mathbf{P}^{(r)\mathrm{T}}\mathbf{H}. \tag{5}$$

Each column of this matrix, denoted $\tilde{\mathbf{y}} = \mathbf{P}^{(r)\mathrm{T}}\mathbf{h}$, represents the projection of a delay vector $\mathbf{h}(t)$ onto the leading r Legendre polynomials. The components of this state vector, $\tilde{\mathbf{y}}(t) = [y_0(t), \dots, y_{r-1}(t)]^{\mathrm{T}}$, are the individual r *Legendre coordinates* [144], which form the input to the neural network encoder.

The dimensionality of $\mathbf{H}$ is commonly reduced through its singular value decomposition (SVD) (e.g., [70, 145]). In practice, we find that projection onto Legendre polynomials yields marginal improvements over SVD of $\mathbf{H}$, particularly in the low-data limit [144]. More importantly, the Legendre polynomials additionally provide a universal, analytic basis in which to represent the dynamics; this property is exploited for comparisons across mice and datasets.

Choice of delay embedding parameters was guided on the basis of autocorrelation time and attractor reconstruction quality (i.e., unfolding the attractor while maximally preserving geometry), following decomposition of $\mathbf{H}$. The number of delay coordinates was guided based upon asymptoting reconstruction performance using a linear regression model. For all main text analyses, the Hankel matrix was constructed with $d = 100$ time delays, with each row separated by $\Delta t = 3$ time steps (equivalently, adopting $\mathbf{H}$ as defined in Eq. 3, we set $d = 300$ and subsampled every third row). As all time series were analyzed at a sampling frequency of 20 Hz, this amounts to a maximum delay window of $100 \times 3 \times .05 = 15$ seconds.

For all modeling, pupil diameter was first shifted in time to accommodate any physiological delay between the widefield signals and the pupil. For each widefield modality, this time shift was selected as the abscissa corresponding to the peak of the cross-correlation function between the pupil and the mean widefield signal. For the leave-one-out pipeline, we used the median time shift across the six mice in the training set.

State-space mappings: variational autoencoder

To interrelate observables through this state-space framework (and thus, through arousal dynamics), we ultimately seek to approximate the target functions

$$\boldsymbol{\psi}_i : \mathbf{h}_i \to \mathbf{z}, \tag{6}$$

$$\boldsymbol{\phi}_j : \mathbf{z} \to \mathbf{y}_j, \tag{7}$$

where $\mathbf{h}_i(t)$ represents the delay vectors (columns of the Hankel matrix constructed from a scalar observable y_i), $\mathbf{z}(t) \in \mathbb{R}^m$ is the low-dimensional representation of the latent arousal state, and $\mathbf{y}_j(t)$ is a second observable. A variety of linear and nonlinear methods can be used to approximate these functions (e.g., see [86]). Our primary results are obtained through a probabilistic modeling architecture based upon the variational autoencoder (VAE) [91, 146].

Briefly, a VAE is a deep generative model comprising a pair of neural networks (i.e., an encoder and decoder) that are jointly trained to map data observations to the mean and variance of a latent distribution (via the encoder), and to map random samples from this latent distribution back to the observation space (decoder). Thus, the VAE assumes that data observations y are taken from a distribution over some latent variable z, such that each data point is treated as a sample $\hat{z}$ from the prior distribution $p_\theta(z)$, typically initialized according to the standard diagonal Gaussian prior, i.e., $p_\theta(z) = \mathcal{N}(z|0, \mathbf{I})$. A variational distribution $q_\varphi(z|y)$ with trainable weights φ is introduced as an approximation to the true (but intractable) posterior distribution $p(z|y)$.

As an autoencoder, $q_\varphi(z|y)$ and $p_\theta(y|z)$ encode observations into a stochastic latent space, and decode from this latent space back to the original observation space. The training goal is to maximize the marginal likelihood of the observables given the latent states z. This problem is made tractable by instead maximizing an evidence lower-bound (ELBO), such that the loss is defined in terms of the negative ELBO:

$$\mathcal{L}_{\varphi,\theta}(y) = -\mathbb{E}_{z\sim q_\varphi(z|y)}\left[\log p_\theta(y|z)\right] + \beta \cdot \mathbb{KL}(q_\varphi(z|y) \| p_\theta(z)), \tag{8}$$

where the first term corresponds to the log-likelihood of the data (reconstruction error), and the Kullback-Leibler (KL) divergence term regularizes the distribution of the latent states $q_\varphi(z|y)$ to be close to that of the prior $p_\theta(z)$. Under Gaussian assumptions, the first term is simply obtained as the mean squared error, i.e., $\|y - \hat{y}\|_2^2$. The KL-term is weighted according to an additional hyperparameter β, which controls the balance between these two losses [147, 148]. Model parameters (φ and θ) can be jointly optimized via stochastic gradient descent through the reparameterization trick [146].

In the present framework, rather than mapping back to the original observation space, we use the decoder $p_\theta(y|z)$ to map from $\mathbf{z}$ to a new observation space. In this way, we use a probabilistic encoder and decoder, respectively, to approximate our target functions $\boldsymbol{\psi}(\mathbf{h})$ (which is partly constituted by the intermediate projection $\mathbf{P}^{(r)}$, via Eq. 5) and $\boldsymbol{\phi}(\mathbf{z})$. Notably, VAE training incorporates stochastic perturbations to the latent representation, thus promoting discovery of a smooth and continuous manifold. Such a representation is desirable in the present framework, such that changes within the latent space (which is based on the delay coordinates) are smoothly mapped to changes within the observation spaces [149]. After training, to generate predicted trajectories, we simply fed the mean output from the encoder (i.e., $q(\mu_z|y)$) through the decoder to (deterministically) generate the maximum a posteriori (MAP) estimate.

Model architecture and hyperparameters

VAE models were trained with the Adam optimizer [150]; see Table 1 for details specific to each dataset. Across datasets, the encoder comprised a single-layer feedforward neural network with tanh activations, which transformed the vector of leading Legendre coordinates $\tilde{\mathbf{y}}$ (Eq. (5)) to (the mean and log variance of) an m-dimensional latent space ($m = 4$ unless otherwise stated). The latent space Gaussian was initialized with a noise scale of $\sigma = .1$ to stabilize training. The decoder from this latent space included a single-layer feedforward neural network consisting of 10 hidden units with tanh activations, followed by a linear readout layer matching the dimensionality of the target signal. During training, the KL divergence weight was gradually annealed to a factor of β over the first $\sim 15\%$ of the training epochs to mitigate posterior collapse [147, 151]. Hyperparameters pertaining to delay embedding and the VAE were primarily selected by examining expressivity and robustness within the training set of a subset

of mice. Within each modeling pipeline/dataset, hyperparameters were kept fixed across all mice and modalities to mitigate overfitting.

Table 1: Neural network hyperparameters across modeling pipelines and datasets.

Hyperparameter	**Leave-one-out**	**Within-mouse**	**Allen mouse**	**Lorenz**
Latent dimensions (z_n)	4	4	2	6
Learning rate (LR)	0.001	0.01	0.001	0.01
Epochs	200	200	30	200
KL weight (β)	0.1	1	0.1	0.1
KL anneal steps	2000	200	2000	2000
Batch size	1000	1000	1000	1000

Model training pipeline

For all main analyses, we implemented an iterative leave-one-out strategy in which candidate models were trained using 10-minute recordings concatenated across six of the seven mice, and subsequently tested on the held-out 10-minute recording obtained from the seventh mouse. To make model training efficient, we employed a randomized singular value decomposition (SVD) approach [68, 152]. Thus, for each iteration, we began by pre-standardizing (to zero mean and unit variance) all datasets ("standardization 1") and aligning in time following delay embedding of the pupil measurements (thus removing the initial time points of the widefield data for which no delay vector of pupil values is available). Next, we lag-adjusted the (delay embedded) pupil and widefield time series from each mouse according to the median lag (across the six training mice) derived from each training mouse's cross-correlation function between their pupil time series and their mean widefield signal. Next, we concatenated the six training datasets and computed the SVD of the concatenated widefield time series ($\mathbf{Y}_{\text{cat}} = \mathbf{U}_{\text{cat}}\mathbf{S}_{\text{cat}}\mathbf{V}_{\text{cat}}^{\text{T}}$) to obtain group-level spatial components $\mathbf{V}_{\text{cat}}$. Data were projected onto the leading 10 components (i.e., the first 10 columns of $\mathbf{V}_{\text{cat}}$) and standardized once more ("standardization 2"), and training proceeded within this reduced subspace. Note that this group-level subspace excludes data from the test mouse, such that $\mathbf{V}_{\text{cat}}$ differs for each iteration of the leave-one-out pipeline.

For the held-out mouse, the pre-standardized pupil diameter (i.e., having undergone "standardization 1" above) was standardized once more using the training set parameters ("standardization 2"), and used to generate a prediction within the group-level 10-dimensional subspace. This prediction was then inverse-transformed and unprojected back into the high-dimensional image space using the training set's standardization and projection parameters. All model evaluations were performed following one final inverse transformation to undo the initial pre-standardization (standardization 1), enabling evaluation in the original dF/F units. The pre-standardization parameters were not predicted from the training data; including the pre-standardization as part of the model would make model performance sensitive to the robustness of the relationship between dF/F values and raw pupil units, which is not scientifically relevant in this context. Finally, any subsequent projections used to evaluate model performance in terms of PCs (as in Fig. 2B) utilized mouse-specific SVD modes (denoted simply as $\mathbf{V}$), rather than the group modes ($\mathbf{V}_{\text{cat}}$) used during training.

In addition to the leave-one-out pipeline, we implemented a separate, within-subject pipeline in which we trained models on the first 5 minutes of data for each mouse and report model performance on the final 3.5 minutes of the 10-minute session. This training pipeline involved prediction of the full-dimensional images, without the intermediate SVD used in the leave-one-out pipeline. All results pertaining to this modeling pipeline are contained in Fig. S9.

Model evaluation

For analyses reporting "Total variance" explained (Figs 2A and 3C, upper), we report the R^2 value over the full 128×128 image, computed across all held-out time points. R^2 was computed as the coefficient

of determination, evaluated over all pixels n and temporal samples t (i.e., the total variance):

$$R^2 = 1 - \frac{\sum_n \sum_t (y_{n,t} - \hat{y}_{n,t})^2}{\sum_n \sum_t (y_{n,t} - \bar{y})^2} \tag{9}$$

where $y_{n,t}$ and $\hat{y}_{n,t}$ are, respectively, the true and predicted values of $\mathbf{y}$ at the n-th pixel and t-th time point, and $\bar{y}$ is the global mean. In practice, this value was computed as the pixel-wise weighted average of R^2 scores, with weights determined by pixel variance (computed using the built-in function *sklearn.metrics.r2_score*, with "multioutput" set to "variance-weighted").

To examine variance explained beyond the first principal component (Figs. 2B and 3C, lower), we computed the (randomized) SVD of the widefield data for each mouse ($\mathbf{Y} = \mathbf{USV}^{\mathrm{T}}$) and, after training, projected the original and reconstructed widefield data onto the top $N = 200$ spatial components excluding the first (i.e, $\mathbf{YV}_{2:N}$) prior to computing R^2 values. The matrices from this SVD were used for all analyses involving projection onto the leading PCs.

For the shuffled control analysis (Fig. S16), we examined total variance explained in widefield data from the test set (as in Fig. 2A), except that for each mouse we swapped the original pupil diameter timecourse with one from each of the other six mice (i.e., we shuffled pupil diameter and widefield calcium data series across mice).

Dynamic mode decomposition

Dynamic mode decomposition (DMD) [153–155] is a data-driven method to dimensionally reduce time series measurements into a superposition of spatiotemporal modes. Given the time series matrix $\mathbf{Y} = \{\mathbf{y}(t_1), \mathbf{y}(t_2), \ldots, \mathbf{y}(t_p)\}$, where $\mathbf{y}(t_i)$ represents the system state vector at the ith time point, the data are split into two matrices, $\mathbf{Y}^{(1)} = \{\mathbf{y}(t_1), \ldots, \mathbf{y}(t_{p-1})\}$ and $\mathbf{Y}^{(2)} = \{\mathbf{y}(t_2), \ldots, \mathbf{y}(t_p)\}$. DMD seeks an approximate linear mapping $\mathbf{A}$ such that

$$\mathbf{Y}^{(2)} \approx \mathbf{A}\mathbf{Y}^{(1)}. \tag{10}$$

Theoretically, the operator $\mathbf{A}$ represents a low-rank approximation to an infinite-dimensional linear operator—namely, the Koopman operator—associated with nonlinear systems [69, 155], motivating use of DMD even in the context of nonlinear dynamics. This low-rank approximation is obtained via SVD of $\mathbf{Y}^{(1)}$ and $\mathbf{Y}^{(2)}$. The measurement data can thus be approximated in terms of the spectral decomposition of $\mathbf{A}$:

$$\mathbf{y}(t) \approx \sum_k^r b_k \mathbf{v}_k e^{\omega_k t} \tag{11}$$

where the eigenvectors $\mathbf{v}_k$ give the dominant spatial modes, the eigenvalues ω_k give the dominant temporal frequencies, and the coefficients b_k determine the amplitudes.

For each mouse, we applied DMD separately to three data matrices: the original widefield calcium images, and the predictions generated through both the "No embedding" model and the "Latent model". Here again, for numerical efficiency, we incorporated randomized SVD as part of the randomized DMD algorithm [156] implemented in *pyDMD* [157, 158].

Following application of DMD, we constructed spatial phase maps from the leading eigenvector of each $\mathbf{A}$ matrix. Specifically, given the leading eigenvector $\mathbf{v}_1 = [v_1, v_2, \ldots, v_n]^{\mathrm{T}} \in \mathbb{C}^n$, we computed the phase at each pixel i as $\theta_i = \arg(v_i)$, where $\arg(\cdot)$ denotes the complex argument of each entry, and v_i represents the value of the eigenvector $\mathbf{v}_1$ at the i-th pixel. Note that the "No embedding" model predictions yielded exclusively real-valued eigenvectors, as these model predictions were fundamentally rank-one (i.e., they reflect linear transformations of the scalar pupil diameter). Accordingly, as the complex argument maps negative real numbers to 0 and positive real numbers to π, phase maps obtained from the "No embedding" model predictions were restricted to these two values.

We used the circular correlation coefficient to quantify the similarity between the phase map obtained from the original dataset and the phase map obtained from either of the two model predictions. Specifically, let θ_i and $\hat{\theta}_i$ represent the phase at spatial location i in the phase maps of the original and

reconstructed datasets. The circular correlation ρ_{circ} between the two phase maps was computed as [159]:

$$\rho_{\text{circ}} = \frac{\sum_{i=1}^{n} \sin\big(\theta_i - \bar{\theta}\big) \sin\Big(\hat{\theta}_i - \bar{\hat{\theta}}\Big)}{\sqrt{\sum_{i=1}^{n} \sin^2(\theta_i - \bar{\theta}) \sum_{i=1}^{n} \sin^2(\hat{\theta}_i - \bar{\hat{\theta}})}}, \tag{12}$$

where $\bar{\theta}$ and $\bar{\hat{\theta}}$ are the mean phase angles (i.e., the mean directions) of the respective maps. The circular correlation is bounded between -1 and 1, where 1 indicates perfect spatial phase alignment between the datasets, and -1 indicates an inverted pattern. However, because the eigenvectors are only determined up to a complex sign, Fig. 2C reports the absolute value this quantity, i.e., $|\rho_{\text{circ}}|$.

Clustering and decoding analysis

We used Gaussian mixture models (GMMs) to assign each (variance-normalized) image frame to one of k clusters (parameterized by the mean and covariance of a corresponding multivariate Gaussian) in an unsupervised fashion. This procedure enables assessment of the spatial specificity of cortical patterns predicted by arousal dynamics, with increasing number of clusters reflecting increased spatial specificity.

Briefly, a GMM models the observed distribution of feature values x as coming from some combination of k Gaussian distributions:

$$p(x|\pi, \mu, \Sigma) = \sum_{i=1}^{k} \pi_i \cdot \mathcal{N}(x|\mu_i, \Sigma_i), \tag{13}$$

where μ_i and Σ_i are the mean and covariance matrix of the ith Gaussian, respectively, and π_i is the probability of x belonging to the ith Gaussian (i.e., the Gaussian weight). We fit the mean, covariance, and weight parameters through the standard expectation-maximization algorithm (EM) as implemented in the GMM package available in *scikit-learn* [160]. This procedure results in a posterior probability for each data point's membership to each of the k Gaussians (also referred to as the "responsibility" of Gaussian i for the data point); (hard) clustering can then be performed by simply assigning each data point to the Gaussian cluster with maximal responsibility.

Images were spatially normalized such that the (brain-masked) image at each time point was set to unit variance (thus yielding a "spatial sign" [161]). Then, for each mouse and each number of clusters k, a GMM with full covariance prior was fit to the normalized image frames to obtain a set of "ground truth" cluster assignments. After training, this GMM was subsequently conditioned on the reconstructed rather than original image frames to obtain maximum a posteriori cluster assignments associated with the "No embedding" or "Latent model" predictions. These cluster assignments were compared to the ground truth assignments, with accuracy computed as the proportion of correctly identified cluster labels.

The preceding algorithm effectively clusters the dimensionally-reduced image frames on the basis of their cosine distance (technically, their L_2-normalized Euclidean distance). Because the first PC tends to explain the vast majority of the variance, this Euclidean-based distance is largely determined by distance along the leading dimension. This analysis choice suffices for the primary purpose of the analysis—namely, to assess correspondence between the original and reconstructed data at the level of individual time points. However, in the context of multivariate data, it is common to also normalize the data *features* prior to clustering, thus equally weighting each (spatial) dimension of the data. To examine this clustering alternative, we performed *feature-wise* normalization of each spatial dimension following projection of the (*sample-wise*) normalized image frames onto the top three PCs. This essentially constitutes a whitening transformation of the image frames, with Euclidean distance in this transformed space effectively approximating the Mahalanobis distance between image frames. Clustering results from this procedure are shown in Fig. S11. For these supplementary analyses, which we also extend to higher cluster numbers, we employ a spherical (rather than full) covariance prior as additional regularization.

Hidden Markov model

We additionally used hidden Markov models (HMMs) to capture the temporal dynamics within the original and reconstructed widefield calcium measurements. Briefly, under a Gaussian HMM, the observations $\{\mathbf{y}(t_1), \mathbf{y}(t_2), \ldots, \mathbf{y}(t_p)\}$ are assumed to be generated by sequence of latent discrete states $z_t \in \{1, \ldots, K\}$, with the full generative model including:

1. A transition matrix $\mathbf{A} \in \mathbb{R}^{K \times K}$, where $A_{ij} = P(z_{t+1} = j | z_t = i)$ represents the probability of transitioning from state i to state j.
2. An initial state distribution, represented as a vector of state probabilities $\boldsymbol{\pi} \in \mathbb{R}^K$ (with $\pi_i = P(z_1 = i)$).
3. Emission distributions, such that each observation $\mathbf{y}_t$ is modeled as a multivariate Gaussian conditioned on the latent state z_t—i.e., $\mathbf{y}_t \sim \mathcal{N}(\boldsymbol{\mu}_z, \boldsymbol{\Sigma}_z)$, where $\boldsymbol{\mu}_z$ and $\boldsymbol{\Sigma}_z$ denote the state-specific mean and covariance, respectively.

The model is thus parameterized by $\boldsymbol{\theta} = \{\mathbf{A}, \boldsymbol{\pi}, \{\boldsymbol{\mu}_z, \boldsymbol{\Sigma}_z\}_{z=1}^K\}$. We estimated these parameters via an EM algorithm (Baum-Welch), as implemented in *hmmlearn* [162].

Prior to HMM fitting, the widefield images were projected onto the three leading PCs and normalized along each dimension (similar to preprocessing for the GMM with whitening above, but without the initial normalization of each individual image frame). HMMs were fit to these data for a range of state numbers ($K = 2, 3, 4, 5, 6$).

To ensure a robust fit, we initialized the state probabilities to be uniform, $\pi_i = \frac{1}{K}$, and initialized the transition matrix with probabilities of .8 along the diagonal (i.e., within-state transitions or "stay" probabilities) and $.2/K$ for all off-diagonal elements (i.e., between-state transitions).

An HMM was first fit to the original data (yielding a "ground truth" model), and then separate HMMs were fit to the reconstructions. In the latter cases, the emission parameters $\{\boldsymbol{\mu}_z, \boldsymbol{\Sigma}_z\}_{z=1}^K$ were fixed to the values obtained from the ground truth model, so that only the transition matrix $\mathbf{A}$ and the initial state probabilities $\boldsymbol{\pi}$ were updated.

For each mouse and each choice of K states, we computed three evaluation metrics:

1. The log-likelihood of the observed data under the fitted HMMs.
2. Decoding accuracy, computed as the agreement between the latent states inferred from the ground truth HMM and those predicted by the HMMs fit to the reconstructed datasets.
3. Transition matrix similarity, which we computed as the row-wise KL-divergence between the ground truth and reconstructed transition matrices:

$$\mathbb{KL}(\mathbf{A}_{\text{true}} \| \mathbf{A}_{\text{model}}) = \frac{1}{K} \sum_{i=1}^{K} \sum_{j=1}^{K} A_{ij}^{\text{true}} \log \frac{A_{ij}^{\text{true}}}{A_{ij}^{\text{model}}}. \tag{14}$$

   For numerical stability in computing KL divergence, transition probabilities below a threshold of $\epsilon = 10^{-6}$ were set to zero and each row was renormalized to maintain stochasticity, i.e.,

$$\sum_{j=1}^{K} A_{ij} = 1 \quad \text{for all } i. \tag{15}$$

Multi-dataset integration

To extend our framework to an independent set of observables, we trained our architecture on concatenated time series from the Allen Institute Brain Observatory dataset ($N = 16$ mice). For visualization purposes, we sought to learn a group-level two-dimensional embedding (i.e., $\mathbf{z} \in \mathbb{R}^2$) based upon delay embedded pupil measurements. After group-level training, we froze the weights of the encoder and retrained the decoders independently for each mouse, enabling individual-specific predictions for the Allen Institute mice from a common (i.e., group-level) latent space. We additionally applied these encoder weights to delay embedded pupil observations from the original "WashU" dataset, once again retraining mouse-specific decoders to reconstruct observables from the common latent space.

Once trained, this procedure results in the generative model $p(\mathbf{y}_i, \mathbf{z})$, which we use to express the posterior probability of each observable $\mathbf{y}_i$ as a function of position within a common 2D latent space $\mathbf{z}$ (note that this formulation does not seek a complete generative model capturing all joint probabilities among the observables). This allows us to systematically evaluate the conditional expectation ($\mathbb{E}(\mathbf{y}_i | \mathbf{z})$) for each

observable over a grid of points in the latent space, which was regularized to be continuous (roughly, an isotropic Gaussian [91]). Fig. 4 represents this expectation, averaged over decoder models trained for each mouse.

Acknowledgements

We wish to thank the Allen Institute for the publicly available data used in this study. We also thank Trevor Voss for assistance with the optical setup. The mouse cartoon in Figs. 1 & 4 is adapted from [163], obtained from SciDraw.io. **Funding:** R.V.R was supported by the Shanahan Family Foundation Fellowship at the Interface of Data and Neuroscience at the Allen Institute and the University of Washington, supported in part by the Allen Institute. This research was additionally supported by the American Heart Association grant 20PRE34990003 (Z.P.R.); National Institutes of Health grants R01NS126326 (A.Q.B), R01NS102870 (A.Q.B), RF1AG07950301 (A.Q.B.), R37NS110699 (J.M.L), R01NS084028 (J.M.L.), and R01NS094692 (J.M.L.); and the National Science Foundation AI Institute in Dynamic Systems grant 2112085 (J.N.K. and S.L.B.).

Data availability statement

The multimodal widefield datasets collected for this study will be deposited in a public repository upon publication. Data from the Allen Brain Observatory can be accessed via the Allen SDK (`https://allensdk.readthedocs.io/en/latest/`).

Code availability statement

All analysis code for this study is publicly available at `https://github.com/ryraut/arousal_dynamics`.

References

1. McCormick, D. A., Nestvogel, D. B. & He, B. J. Neuromodulation of Brain State and Behavior. *Annual Review of Neuroscience* **43,** 391–415. ISSN: 0147-006X, 1545-4126. doi:10.1146/annurev-neuro-100219-105424 (2020).

2. Buzsáki, G. *The Brain from Inside Out* Type: Book (Oxford University Press, New York, NY, USA, 2019).

3. Raichle, M. E. The restless brain: how intrinsic activity organizes brain function. *Philosophical Transactions of the Royal Society B: Biological Sciences* **370,** 20140172. ISSN: 0962-8436, 1471-2970. doi:10.1098/rstb.2014.0172 (2015).

4. Engel, A. K., Gerloff, C., Hilgetag, C. C. & Nolte, G. Intrinsic Coupling Modes: Multiscale Interactions in Ongoing Brain Activity. *Neuron* **80,** 867–886. ISSN: 08966273. doi:10.1016/j.neuron.2013.09.038 (2013).

5. Nir, Y. & de Lecea, L. Sleep and vigilance states: Embracing spatiotemporal dynamics. *Neuron.* ISSN: 0896-6273. doi:10.1016/j.neuron.2023.04.012 (2023).

6. Greene, A. S., Horien, C., Barson, D., Scheinost, D. & Constable, R. T. Why is everyone talking about brain state? *Trends in Neurosciences* **0.** Publisher: Elsevier. ISSN: 0166-2236, 1878-108X. doi:10.1016/j.tins.2023.04.001 (2023).

7. Stringer, C. *et al.* Spontaneous behaviors drive multidimensional, brainwide activity. *Science* **364,** eaav7893. ISSN: 0036-8075, 1095-9203. doi:10.1126/science.aav7893 (2019).

8. Shimaoka, D., Harris, K. D. & Carandini, M. Effects of Arousal on Mouse Sensory Cortex Depend on Modality. *Cell Reports* **22,** 3160–3167. ISSN: 22111247. doi:10.1016/j.celrep.2018.02.092 (2018).

9. Musall, S., Kaufman, M. T., Juavinett, A. L., Gluf, S. & Churchland, A. K. Single-trial neural dynamics are dominated by richly varied movements. *Nature Neuroscience* **22,** 1677–1686. ISSN: 1097-6256, 1546-1726. doi:10.1038/s41593-019-0502-4 (2019).

10. Kaplan, H. S. & Zimmer, M. Brain-wide representations of ongoing behavior: a universal principle? *Current Opinion in Neurobiology* **64,** 60–69. ISSN: 09594388. doi:10.1016/j.conb.2020.02.008 (2020).

11. Salkoff, D. B., Zagha, E., McCarthy, E. & McCormick, D. A. Movement and Performance Explain Widespread Cortical Activity in a Visual Detection Task. *Cerebral Cortex* **30,** 421–437. ISSN: 1047-3211, 1460-2199. doi:10.1093/cercor/bhz206 (2020).

12. Syeda, A. *et al.* Facemap: a framework for modeling neural activity based on orofacial tracking. *Nature Neuroscience* **27.** Publisher: Nature Publishing Group, 187–195. ISSN: 1546-1726. doi:10.1038/s41593-023-01490-6 (2024).

13. Liu, X., Leopold, D. A. & Yang, Y. Single-neuron firing cascades underlie global spontaneous brain events. *Proceedings of the National Academy of Sciences* **118,** e2105395118. ISSN: 0027-8424, 1091-6490. doi:10.1073/pnas.2105395118 (2021).

14. Crombie, D., Spacek, M. A., Leibold, C. & Busse, L. Spiking activity in the visual thalamus is coupled to pupil dynamics across temporal scales. *PLOS Biology* **22.** Publisher: Public Library of Science, e3002614. ISSN: 1545-7885. doi:10.1371/journal.pbio.3002614 (2024).

15. Vickers, E. D. & McCormick, D. A. Pan-cortical 2-photon mesoscopic imaging and neurobehavioral alignment in awake, behaving mice. *eLife* **13.** Publisher: eLife Sciences Publications Limited. doi:10.7554/eLife.94167.2 (2024).

16. Drew, P. J., Winder, A. T. & Zhang, Q. Twitches, Blinks, and Fidgets: Important Generators of Ongoing Neural Activity. *The Neuroscientist* **25,** 298–313. ISSN: 1073-8584, 1089-4098. doi:10.1177/1073858418805427 (2019).

17. Sadeh, S. & Clopath, C. Contribution of behavioural variability to representational drift. *eLife* **11.** Publisher: eLife Sciences Publications, Ltd, e77907. ISSN: 2050-084X. doi:10.7554/eLife.77907 (2022).

18. Zagha, E. *et al.* The Importance of Accounting for Movement When Relating Neuronal Activity to Sensory and Cognitive Processes. *The Journal of Neuroscience* **42,** 1375–1382. ISSN: 0270-6474, 1529-2401. doi:10.1523/JNEUROSCI.1919-21.2021 (2022).

19. Collins, L., Boddington, L., Steffan, P. J. & McCormick, D. Vagus nerve stimulation induces widespread cortical and behavioral activation. *Current Biology* **31,** 2088–2098.e3. ISSN: 09609822. doi:10.1016/j.cub.2021.02.049 (2021).

20. Lohani, S. *et al.* Spatiotemporally heterogeneous coordination of cholinergic and neocortical activity. *Nature Neuroscience* **25.** Number: 12 Publisher: Nature Publishing Group, 1706–1713. ISSN: 1546-1726. doi:10.1038/s41593-022-01202-6 (2022).

21. Saper, C. B. & Fuller, P. M. Wake–sleep circuitry: an overview. *Current Opinion in Neurobiology* **44,** 186–192. ISSN: 09594388. doi:10.1016/j.conb.2017.03.021 (2017).

22. Scammell, T. E., Arrigoni, E. & Lipton, J. O. Neural Circuitry of Wakefulness and Sleep. *Neuron* **93,** 747–765. ISSN: 08966273. doi:10.1016/j.neuron.2017.01.014 (2017).

23. Bechtel, W. & Abrahamsen, A. in *Philosophy of Complex Systems* (ed Hooker, C.) 257–285 (North-Holland, Amsterdam, 2011). doi:10.1016/B978-0-444-52076-0.50009-2.

24. Duffy, E. The psychological significance of the concept of "arousal" or "activation." *Psychological Review* **64.** Place: US Publisher: American Psychological Association, 265–275. ISSN: 1939-1471. doi:10.1037/h0048837 (1957).

25. Kahneman, D. *Attention and effort* First Edition. 246 pp. ISBN: 978-0-13-050518-7 (Prentice-Hall, Englewood Cliffs, N.J, 1973).

26. Andrew, R. J. Arousal and the Causation of Behaviour. *Behaviour* **51.** Publisher: Brill, 135–165. ISSN: 0005-7959 (1974).

27. Calderon, D., Kilinc, M., Maritan, A., Banavar, J. & Pfaff, D. Generalized CNS arousal: An elementary force within the vertebrate nervous system. *Neuroscience & Biobehavioral Reviews* **68,** 167–176. ISSN: 01497634. doi:10.1016/j.neubiorev.2016.05.014 (2016).

28. Dampney, R. A. L. Central neural control of the cardiovascular system: current perspectives. *Advances in Physiology Education* **40.** Publisher: American Physiological Society, 283–296. ISSN: 1043-4046. doi:10.1152/advan.00027.2016 (2016).

29. Rasmussen, R., O'Donnell, J., Ding, F. & Nedergaard, M. Interstitial ions: A key regulator of state-dependent neural activity? *Prog Neurobiol* **193.** Type: Journal Article, 101802. ISSN: 1873-5118. doi:10.1016/j.pneurobio.2020.101802 (2020).

30. Roeder, T. The control of metabolic traits by octopamine and tyramine in invertebrates. *Journal of Experimental Biology* **223,** jeb194282. ISSN: 0022-0949. doi:10.1242/jeb.194282 (2020).

31. Martin, C. G., He, B. J. & Chang, C. State-related neural influences on fMRI connectivity estimation. *NeuroImage* **244,** 118590. ISSN: 1053-8119. doi:10.1016/j.neuroimage.2021.118590 (2021).

32. McGinley, M. J. *et al.* Waking State: Rapid Variations Modulate Neural and Behavioral Responses. *Neuron* **87,** 1143–1161. ISSN: 08966273. doi:10.1016/j.neuron.2015.09.012 (2015).

33. Mann, K., Deny, S., Ganguli, S. & Clandinin, T. R. Coupling of activity, metabolism and behaviour across the Drosophila brain. *Nature* **593,** 244–248. ISSN: 0028-0836, 1476-4687. doi:10.1038/s41586-021-03497-0 (2021).

34. Grujic, N., Polania, R. & Burdakov, D. Neurobehavioral meaning of pupil size. *Neuron* **0.** Publisher: Elsevier. ISSN: 0896-6273. doi:10.1016/j.neuron.2024.05.029 (2024).

35. Poulet, J. F. A. & Crochet, S. The Cortical States of Wakefulness. *Frontiers in Systems Neuroscience* **12,** 64. ISSN: 1662-5137. doi:10.3389/fnsys.2018.00064 (2019).

36. Fox, M. D. & Raichle, M. E. Spontaneous fluctuations in brain activity observed with functional magnetic resonance imaging. *Nature Reviews Neuroscience* **8,** 700–711. ISSN: 1471-003X, 1471-0048. doi:10.1038/nrn2201 (2007).

37. Pais-Roldán, P. *et al.* Contribution of animal models toward understanding resting state functional connectivity. *NeuroImage* **245,** 118630. ISSN: 1053-8119. doi:10.1016/j.neuroimage.2021.118630 (2021).

38. Grandjean, J. *et al.* A consensus protocol for functional connectivity analysis in the rat brain. *Nature Neuroscience* **26.** Number: 4 Publisher: Nature Publishing Group, 673–681. ISSN: 1546-1726. doi:10.1038/s41593-023-01286-8 (2023).

39. Cabral, J., Kringelbach, M. L. & Deco, G. Functional connectivity dynamically evolves on multiple time-scales over a static structural connectome: Models and mechanisms. *NeuroImage* **160,** 84–96. ISSN: 10538119. doi:10.1016/j.neuroimage.2017.03.045 (2017).

40. Reid, A. T. *et al.* Advancing functional connectivity research from association to causation. *Nature Neuroscience* **22,** 1751–1760. ISSN: 1097-6256, 1546-1726. doi:10.1038/s41593-019-0510-4 (2019).

41. Suárez, L. E., Markello, R. D., Betzel, R. F. & Misic, B. Linking Structure and Function in Macroscale Brain Networks. *Trends in Cognitive Sciences* **24,** 302–315. ISSN: 13646613. doi:10.1016/j.tics.2020.01.008 (2020).

42. Gu, Y., Han, F. & Liu, X. Arousal Contributions to Resting-State fMRI Connectivity and Dynamics. *Frontiers in Neuroscience* **13,** 1190. ISSN: 1662-453X. doi:10.3389/fnins.2019.01190 (2019).

43. Shine, J. M. Neuromodulatory Influences on Integration and Segregation in the Brain. *Trends in Cognitive Sciences* **23,** 572–583. ISSN: 1364-6613. doi:10.1016/j.tics.2019.04.002 (2019).

44. Gutierrez-Barragan, D. *et al.* Unique spatiotemporal fMRI dynamics in the awake mouse brain. *Current Biology* **32,** 631–644.e6. ISSN: 09609822. doi:10.1016/j.cub.2021.12.015 (2022).

45. West, S. L. *et al.* Wide-Field Calcium Imaging of Dynamic Cortical Networks during Locomotion. *Cerebral Cortex* **32,** 2668–2687. ISSN: 1047-3211. doi:10.1093/cercor/bhab373 (2022).

46. Shahsavarani, S. *et al.* Cortex-wide neural dynamics predict behavioral states and provide a neural basis for resting-state dynamic functional connectivity. *Cell Reports* **42.** Publisher: Elsevier. ISSN: 2211-1247. doi:10.1016/j.celrep.2023.112527 (2023).

47. Benisty, H. *et al.* Rapid fluctuations in functional connectivity of cortical networks encode spontaneous behavior. *Nature Neuroscience.* Publisher: Nature Publishing Group, 1–11. ISSN: 1546-1726. doi:10.1038/s41593-023-01498-y (2023).

48. Power, J. D., Plitt, M., Laumann, T. O. & Martin, A. Sources and implications of whole-brain fMRI signals in humans. *NeuroImage* **146,** 609–625. ISSN: 10538119. doi:10.1016/j.neuroimage.2016.09.038 (2017).

49. Turchi, J. *et al.* The Basal Forebrain Regulates Global Resting-State fMRI Fluctuations. *Neuron* **97,** 940–952.e4. ISSN: 08966273. doi:10.1016/j.neuron.2018.01.032 (2018).

50. Özbay, P. S. *et al.* Sympathetic activity contributes to the fMRI signal. *Communications Biology* **2,** 421. ISSN: 2399-3642. doi:10.1038/s42003-019-0659-0 (2019).

51. Bolt, T. S. *et al. A Unified Physiological Process Links Global Patterns of Functional MRI, Respiratory Activity, and Autonomic Signaling* Pages: 2023.01.19.524818 Section: New Results. 2023. doi:10.1101/2023.01.19.524818.

52. Matsui, T., Murakami, T. & Ohki, K. Transient neuronal coactivations embedded in globally propagating waves underlie resting-state functional connectivity. *Proceedings of the National Academy of Sciences* **113,** 6556–6561. ISSN: 0027-8424, 1091-6490. doi:10.1073/pnas.1521299113 (2016).

53. Mitra, A. *et al.* Spontaneous Infra-slow Brain Activity Has Unique Spatiotemporal Dynamics and Laminar Structure. *Neuron* **98,** 297–305.e6. ISSN: 08966273. doi:10.1016/j.neuron.2018.03.015 (2018).

54. Chen, J. E. *et al.* Resting-state "physiological networks". *NeuroImage* **213,** 116707. ISSN: 10538119. doi:10.1016/j.neuroimage.2020.116707 (2020).

55. Yousefi, B. & Keilholz, S. Propagating patterns of intrinsic activity along macroscale gradients coordinate functional connections across the whole brain. *NeuroImage* **231,** 117827. ISSN: 10538119. doi:10.1016/j.neuroimage.2021.117827 (2021).

56. Bolt, T. *et al.* A parsimonious description of global functional brain organization in three spatiotemporal patterns. *Nature Neuroscience.* Publisher: Nature Publishing Group, 1–11. ISSN: 1546-1726. doi:10.1038/s41593-022-01118-1 (2022).

57. Pang, J. C. *et al.* Geometric constraints on human brain function. *Nature.* Publisher: Nature Publishing Group, 1–9. ISSN: 1476-4687. doi:10.1038/s41586-023-06098-1 (2023).

58. Yellin, D., Berkovich-Ohana, A. & Malach, R. Coupling between pupil fluctuations and resting-state fMRI uncovers a slow build-up of antagonistic responses in the human cortex. *NeuroImage* **106,** 414–427. ISSN: 10538119. doi:10.1016/j.neuroimage.2014.11.034 (2015).

59. Hindriks, R., R, M., N, G. & G, D. Latency analysis of resting-state BOLD-fMRI reveals traveling waves in visual cortex linking task-positive and task-negative networks. *NeuroImage* **200,** 259–274. ISSN: 10538119. doi:10.1016/j.neuroimage.2019.06.007 (2019).

60. Shine, J. M. *et al.* Human cognition involves the dynamic integration of neural activity and neuromodulatory systems. *Nature Neuroscience* **22,** 289–296. ISSN: 1097-6256, 1546-1726. doi:10.1038/s41593-018-0312-0 (2019).

61. Brown, J. A., Lee, A. J., Pasquini, L. & Seeley, W. W. A dynamic gradient architecture generates brain activity states. *NeuroImage* **261,** 119526. ISSN: 1053-8119. doi:10.1016/j.neuroimage.2022.119526 (2022).

62. Gutierrez-Barragan, D., Ramirez, J. S. B., Panzeri, S., Xu, T. & Gozzi, A. Evolutionarily conserved fMRI network dynamics in the mouse, macaque, and human brain. *Nature Communications* **15.** Publisher: Nature Publishing Group, 8518. ISSN: 2041-1723. doi:10.1038/s41467-024-52721-8 (2024).

63. Raut, R. V. *et al.* Global waves synchronize the brain's functional systems with fluctuating arousal. *Science Advances* **7,** eabf2709. ISSN: 2375-2548. doi:10.1126/sciadv.abf2709 (2021).

64. Strogatz, S. H. *Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry, and Engineering* 2nd ed. 532 pp. ISBN: 978-0-429-49256-3. doi:10.1201/9780429492563 (CRC Press, Boca Raton, 2019).

65. Takens, F. Detecting strange attractors in turbulence. *Lecture Notes in Mathematics* **898,** 366–381 (1981).

66. Stark, J., Broomhead, D. S., Davies, M. E. & Huke, J. Takens embedding theorems for forced and stochastic systems. *Nonlinear Analysis: Theory, Methods & Applications. Proceedings of the Second World Congress of Nonlinear Analysts* **30,** 5303–5314. ISSN: 0362-546X. doi:10.1016/S0362-546X(96)00149-6 (1997).

67. Sugihara, G. *et al.* Detecting Causality in Complex Ecosystems. *Science* **338,** 496–500. ISSN: 0036-8075, 1095-9203. doi:10.1126/science.1227079 (2012).

68. Brunton, S. L. & Kutz, J. N. *Data-Driven Science and Engineering: Machine Learning, Dynamical Systems, and Control* 2nd (Cambridge University Press, 2022).

69. Brunton, S. L., Budišić, M., Kaiser, E. & Kutz, J. N. Modern Koopman Theory for Dynamical Systems. *SIAM Review* **64,** 229–340. ISSN: 0036-1445, 1095-7200. doi:10.1137/21M1401243 (2022).

70. Brunton, S. L., Brunton, B. W., Proctor, J. L., Kaiser, E. & Kutz, J. N. Chaos as an intermittently forced linear system. *Nature Communications* **8,** 19. ISSN: 2041-1723. doi:10.1038/s41467-017-00030-8 (2017).

71. Wang, X. *et al.* Spatiotemporal relationships between neuronal, metabolic, and hemodynamic signals in the awake and anesthetized mouse brain. *Cell Reports* **43,** 114723. ISSN: 2211-1247. doi:10.1016/j.celrep.2024.114723 (2024).

72. Reitman, M. E. *et al.* Norepinephrine links astrocytic activity to regulation of cortical state. *Nature Neuroscience* **26.** Number: 4 Publisher: Nature Publishing Group, 579–593. ISSN: 1546-1726. doi:10.1038/s41593-023-01284-w (2023).

73. Malmo, R. B. Activation: A neuropsychological dimension. *Psychological Review* **66.** Place: US Publisher: American Psychological Association, 367–386. ISSN: 1939-1471. doi:10.1037/h0047858 (1959).

74. Thayer, R. E. Toward a psychological theory of multidimensional activation (arousal). *Motivation and Emotion* **2,** 1–34. ISSN: 1573-6644. doi:10.1007/BF00992729 (1978).

75. Dyer, E. L. & Kording, K. Why the simplest explanation isn't always the best. *Proceedings of the National Academy of Sciences* **120.** Publisher: Proceedings of the National Academy of Sciences, e2319169120. doi:10.1073/pnas.2319169120 (2023).

76. Meilă, M. & Zhang, H. Manifold Learning: What, How, and Why. *Annual Review of Statistics and Its Application* **11.** Publisher: Annual Reviews, 393–417. ISSN: 2326-8298, 2326-831X. doi:10.1146/annurev-statistics-040522-115238 (Volume 11, 2024 2024).

77. Gould, S. J. in *The mismeasure of man* 1st ed (Norton, New York, 1981). ISBN: 978-0-393-01489-1.

78. Bich, L., Mossio, M., Ruiz-Mirazo, K. & Moreno, A. Biological regulation: controlling the system from within. *Biology & Philosophy* **31,** 237–265. ISSN: 0169-3867, 1572-8404. doi:10.1007/s10539-015-9497-8 (2016).

79. Berntson, G. G., Cacioppo, J. T. & Quigley, K. S. Autonomic determinism: The modes of autonomic control, the doctrine of autonomic space, and the laws of autonomic constraint. *Psychological Review* **98,** 459–487. ISSN: 1939-1471, 0033-295X. doi:10.1037/0033-295X.98.4.459 (1991).

80. Schöner, G. & Kelso, J. A. S. Dynamic Pattern Generation in Behavioral and Neural Systems. *Science* **239,** 1513–1520. ISSN: 0036-8075, 1095-9203. doi:10.1126/science.3281253 (1988).

81. Givon, D., Kupferman, R. & Stuart, A. Extracting macroscopic dynamics: model problems and algorithms. *Nonlinearity* **17,** R55–R127. ISSN: 0951-7715, 1361-6544. doi:10.1088/0951-7715/17/6/R01 (2004).

82. Haken, H. *Information and Self-Organization* ISBN: 978-3-540-33021-9. doi:10.1007/3-540-33023-2 (Springer Berlin Heidelberg, 2006).

83. Bertalanffy, L. v. *General System Theory: Foundations, Development, Applications* Second Printing edition. 289 pp. (George Braziller, 1969).

84. Macke, J. H. *et al. Empirical models of spiking in neural populations* in *Advances in Neural Information Processing Systems* **24** (Curran Associates, Inc., 2011).

85. Mezić, I. Spectral properties of dynamical systems, model reduction and decompositions. *Nonlinear Dynamics* **41,** 309–325 (2005).

86. Abarbanel, H. D. I., Carroll, T. A., Pecora, L. M., Sidorowich, J. J. & Tsimring, L. S. Predicting physical variables in time-delay embedding. *Physical Review E* **49.** Publisher: American Physical Society, 1840–1853. doi:10.1103/PhysRevE.49.1840 (1994).

87. Rulkov, N. F., Sushchik, M. M., Tsimring, L. S. & Abarbanel, H. D. I. Generalized synchronization of chaos in directionally coupled chaotic systems. *Physical Review E* **51,** 980–994. ISSN: 1063-651X, 1095-3787. doi:10.1103/PhysRevE.51.980 (1995).

88. Aladjalova, N. A. Infra-slow rhythmic oscillations of the steady potential of the cerebral cortex. *Nature* **179,** 957–959. ISSN: 0028-0836. doi:10.1038/179957a0 (1957).

89. Palva, J. M. & Palva, S. Infra-slow fluctuations in electrophysiological recordings, blood-oxygenation-level-dependent signals, and psychophysical time series. *NeuroImage* **62,** 2201–2211. ISSN: 10538119. doi:10.1016/j.neuroimage.2012.02.060 (2012).

90. Okun, M., Steinmetz, N. A., Lak, A., Dervinis, M. & Harris, K. D. Distinct Structure of Cortical Population Activity on Fast and Infraslow Timescales. *Cerebral Cortex* **29,** 2196–2210. ISSN: 1047-3211, 1460-2199. doi:10.1093/cercor/bhz023 (2019).

91. Kingma, D. P. & Welling, M. An Introduction to Variational Autoencoders. *Foundations and Trends® in Machine Learning* **12,** 307–392. ISSN: 1935-8237, 1935-8245. doi:10.1561/2200000056. arXiv: 1906.02691[cs,stat] (2019).

92. Barson, D. *et al.* Simultaneous mesoscopic and two-photon imaging of neuronal activity in cortical circuits. *Nature Methods* **17,** 107–113. ISSN: 1548-7091, 1548-7105. doi:10.1038/s41592-019-0625-2 (2020).

93. Meyer-Baese, L. *et al.* Cortical Networks Relating to Arousal Are Differentially Coupled to Neural Activity and Hemodynamics. *Journal of Neuroscience* **44.** Publisher: Society for Neuroscience Section: Research Articles. ISSN: 0270-6474, 1529-2401. doi:10.1523/JNEUROSCI.0298-23.2024 (2024).

94. Reimer, J. *et al.* Pupil fluctuations track rapid changes in adrenergic and cholinergic activity in cortex. *Nature Communications* **7,** 13289. ISSN: 2041-1723. doi:10.1038/ncomms13289 (2016).

95. Collins, L., Francis, J., Emanuel, B. & McCormick, D. A. Cholinergic and noradrenergic axonal activity contains a behavioral-state signal that is coordinated across the dorsal cortex. *eLife* **12** (eds Nelson, S. B., Huguenard, J. R. & London, M.) Publisher: eLife Sciences Publications, Ltd, e81826. ISSN: 2050-084X. doi:10.7554/eLife.81826 (2023).

96. Vanni, M. P., Chan, A. W., Balbi, M., Silasi, G. & Murphy, T. H. Mesoscale Mapping of Mouse Cortex Reveals Frequency-Dependent Cycling between Distinct Macroscale Functional Modules. *The Journal of Neuroscience* **37,** 7513–7533. ISSN: 0270-6474, 1529-2401. doi:10.1523/JNEUROSCI.3560-16.2017 (2017).

97. Majeed, W. *et al.* Spatiotemporal dynamics of low frequency BOLD fluctuations in rats and humans. *NeuroImage* **54,** 1140–1150. ISSN: 10538119. doi:10.1016/j.neuroimage.2010.08.030 (2011).

98. Zuend, M. *et al.* Arousal-induced cortical activity triggers lactate release from astrocytes. *Nature Metabolism* **2,** 179–191. ISSN: 2522-5812. doi:10.1038/s42255-020-0170-4 (2020).

99. Schaeffer, S. & Iadecola, C. Revisiting the neurovascular unit. *Nature Neuroscience* **24,** 1198–1209. ISSN: 1097-6256, 1546-1726. doi:10.1038/s41593-021-00904-7 (2021).

100. Shvets-Ténéta-Gurii, T. B., Troshin, G. I. & Dubinin, A. G. Changes in the redox potential of the rabbit cerebral cortex accompanying episodes of ECoG arousal during slow-wave sleep. *Neuroscience and Behavioral Physiology* **38,** 63–70. ISSN: 1573-899X. doi:10.1007/s11055-008-0009-z (2008).

101. Kolenc, O. I. & Quinn, K. P. Evaluating Cell Metabolism Through Autofluorescence Imaging of NAD(P)H and FAD. *Antioxidants & Redox Signaling* **30,** 875–889. ISSN: 1523-0864, 1557-7716. doi:10.1089/ars.2017.7451 (2019).

102. Wang, C.-A. & Munoz, D. P. Coordination of Pupil and Saccade Responses by the Superior Colliculus. *Journal of Cognitive Neuroscience* **33,** 919–932. ISSN: 0898-929X. doi:10.1162/jocn_a_01688 (2021).

103. Dabagia, M., Kording, K. P. & Dyer, E. L. Aligning latent representations of neural activity. *Nature Biomedical Engineering.* Publisher: Nature Publishing Group, 1–7. ISSN: 2157-846X. doi:10.1038/s41551-022-00962-7 (2022).

104. Siegle, J. H. *et al.* Survey of spiking in the mouse visual system reveals functional hierarchy. *Nature.* ISSN: 0028-0836, 1476-4687. doi:10.1038/s41586-020-03171-x (2021).

105. Jun, J. J. *et al.* Fully integrated silicon probes for high-density recording of neural activity. *Nature* **551,** 232–236. ISSN: 0028-0836, 1476-4687. doi:10.1038/nature24636 (2017).

106. Nestvogel, D. B. & McCormick, D. A. Visual thalamocortical mechanisms of waking state-dependent activity and alpha oscillations. *Neuron* **110,** 120–138.e4. ISSN: 08966273. doi:10.1016/j.neuron.2021.10.005 (2022).

107. Brunton, S. L., Proctor, J. L. & Kutz, J. N. Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proceedings of the National Academy of Sciences* **113,** 3932–3937. ISSN: 0027-8424, 1091-6490. doi:10.1073/pnas.1517384113 (2016).

108. Sirota, A., Csicsvari, J., Buhl, D. & Buzsáki, G. Communication between neocortex and hippocampus during sleep in rodents. *Proceedings of the National Academy of Sciences* **100,** 2065–2069. ISSN: 0027-8424, 1091-6490. doi:10.1073/pnas.0437938100 (2003).

109. McGinley, M. J., David, S. V. & McCormick, D. A. Cortical Membrane Potential Signature of Optimal States for Sensory Signal Detection. *Neuron* **87,** 179–192. ISSN: 08966273. doi:10.1016/j.neuron.2015.05.038 (2015).

110. Reimer, J. *et al.* Pupil Fluctuations Track Fast Switching of Cortical States during Quiet Wakefulness. *Neuron* **84,** 355–362. ISSN: 08966273. doi:10.1016/j.neuron.2014.09.033 (2014).

111. Liu, X. *et al.* Arousal transitions in sleep, wakefulness, and anesthesia are characterized by an orderly sequence of cortical events. *NeuroImage* **116,** 222–231. ISSN: 10538119. doi:10.1016/j.neuroimage.2015.04.003 (2015).

112. Varela, F. J., Thompson, E. & Rosch, E. *The embodied mind : cognitive science and human experience* Type: Book. xx, 308 p. ISBN: 978-0-262-22042-2 (MIT Press, Cambridge, Mass., 1991).

113. Kay, K. & Frank, L. M. Three brain states in the hippocampus and cortex. *Hippocampus* **29,** 184–238. ISSN: 10509631. doi:10.1002/hipo.22956 (2019).

114. Franke, K. *et al.* State-dependent pupil dilation rapidly shifts visual feature selectivity. *Nature.* Publisher: Nature Publishing Group, 1–7. ISSN: 1476-4687. doi:10.1038/s41586-022-05270-3 (2022).

115. Gervasoni, D. Global Forebrain Dynamics Predict Rat Behavioral States and Their Transitions. *Journal of Neuroscience* **24,** 11137–11147. ISSN: 0270-6474, 1529-2401. doi:10.1523/JNEUROSCI.3524-04.2004 (2004).

116. Vidaurre, D., Smith, S. M. & Woolrich, M. W. Brain network dynamics are hierarchically organized in time. *Proceedings of the National Academy of Sciences* **114,** 12827–12832. ISSN: 0027-8424, 1091-6490. doi:10.1073/pnas.1705120114 (2017).

117. Schneider, M. *et al.* Spontaneous pupil dilations during the resting state are associated with activation of the salience network. *NeuroImage* **139,** 189–201. ISSN: 10538119. doi:10.1016/j.neuroimage.2016.06.011 (2016).

118. Schreiber, T. & Schmitz, A. Surrogate time series. *Physica D: Nonlinear Phenomena* **142,** 346–382. ISSN: 01672789. doi:10.1016/S0167-2789(00)00043-9 (2000).

119. Sadaghiani, S., Brookes, M. J. & Baillet, S. Connectomics of human electrophysiology. *NeuroImage* **247,** 118788. ISSN: 10538119. doi:10.1016/j.neuroimage.2021.118788 (2022).

120. Chan, A. W., Mohajerani, M. H., LeDue, J. M., Wang, Y. T. & Murphy, T. H. Mesoscale infraslow spontaneous membrane potential fluctuations recapitulate high-frequency activity cortical motifs. *Nature Communications* **6,** 7738. ISSN: 2041-1723. doi:10.1038/ncomms8738 (2015).

121. Ye, Z. *et al. Brain-wide topographic coordination of traveling spiral waves* Pages: 2023.12.07.570517 Section: New Results. 2023. doi:10.1101/2023.12.07.570517.

122. Raichle, M. E., Hartman, B. K., Eichling, J. O. & Sharpe, L. G. Central noradrenergic regulation of cerebral blood flow and vascular permeability. *Proceedings of the National Academy of Sciences* **72.** Publisher: Proceedings of the National Academy of Sciences, 3726–3730. doi:10.1073/pnas.72.9.3726 (1975).

123. Harik, S. I., LaManna, J. C., Light, A. I. & Rosenthal, M. Cerebral Norepinephrine: Influence on Cortical Oxidative Metabolism in Situ. *Science* **206,** 69–71. ISSN: 0036-8075, 1095-9203. doi:10.1126/science.482927 (1979).

124. Lou, H. C., Edvinsson, L. & MacKenzie, E. T. The concept of coupling blood flow to brain function: Revision required? *Annals of Neurology* **22,** 289–297. ISSN: 0364-5134, 1531-8249. doi:10.1002/ana.410220302 (1987).

125. Cohen, Z., Molinatti, G. & Hamel, E. Astroglial and Vascular Interactions of Noradrenaline Terminals in the Rat Cerebral Cortex. *Journal of Cerebral Blood Flow & Metabolism* **17.** Publisher:

SAGE Publications Ltd STM, 894–904. ISSN: 0271-678X. doi:10.1097/00004647-199708000-00008 (1997).

126. Fu, Y. *et al.* A Cortical Circuit for Gain Control by Behavioral State. *Cell* **156,** 1139–1152. ISSN: 00928674. doi:10.1016/j.cell.2014.01.050 (2014).

127. Magistretti, P. & Morrison, J. Noradrenaline- and vasoactive intestinal peptide-containing neuronal systems in neocortex: Functional convergence with contrasting morphology. *Neuroscience* **24,** 367–378. ISSN: 03064522. doi:10.1016/0306-4522(88)90338-7 (1988).

128. Mateo, C., Knutsen, P. M., Tsai, P. S., Shih, A. Y. & Kleinfeld, D. Entrainment of Arteriole Vasomotor Fluctuations by Neural Activity Is a Basis of Blood-Oxygenation-Level-Dependent "Resting-State" Connectivity. *Neuron* **96,** 936–948.e3. ISSN: 08966273. doi:10.1016/j.neuron.2017.10.012 (2017).

129. Zhang, Q. *et al.* Cerebral oxygenation during locomotion is modulated by respiration. *Nature Communications* **10,** 5515. ISSN: 2041-1723. doi:10.1038/s41467-019-13523-5 (2019).

130. Strauch, C., Wang, C.-A., Einhäuser, W., Van der Stigchel, S. & Naber, M. Pupillometry as an integrated readout of distinct attentional networks. *Trends in Neurosciences,* S0166223622000972. ISSN: 01662236. doi:10.1016/j.tins.2022.05.003 (2022).

131. Beam, E., Potts, C., Poldrack, R. A. & Etkin, A. A data-driven framework for mapping domains of human neurobiology. *Nature Neuroscience* **24,** 1733–1744. ISSN: 1097-6256, 1546-1726. doi:10.1038/s41593-021-00948-9 (2021).

132. Dana, H. *et al.* Thy1 transgenic mice expressing the red fluorescent calcium indicator jRGECO1a for neuronal population imaging in vivo. *PLOS ONE* **13.** Publisher: Public Library of Science, e0205444. ISSN: 1932-6203. doi:10.1371/journal.pone.0205444 (2018).

133. Rosenthal, Z. P. *et al.* Local Perturbations of Cortical Excitability Propagate Differentially Through Large-Scale Functional Networks. *Cerebral Cortex* **30,** 3352–3369. ISSN: 1047-3211, 1460-2199. doi:10.1093/cercor/bhz314 (2020).

134. Padawer-Curry, J. A. *et al. Psychedelic 5-HT2A receptor agonism: neuronal signatures and altered neurovascular coupling* Pages: 2023.09.23.559145 Section: New Results. 2023. doi:10.1101/2023.09.23.559145.

135. White, B. R. *et al.* Imaging of Functional Connectivity in the Mouse Brain. *PLoS ONE* **6** (ed Sporns, O.) e16322. ISSN: 1932-6203. doi:10.1371/journal.pone.0016322 (2011).

136. Ma, Y. *et al.* Wide-field optical mapping of neural activity and brain haemodynamics: considerations and novel approaches. *Philosophical Transactions of the Royal Society B: Biological Sciences* **371,** 20150360. ISSN: 0962-8436, 1471-2970. doi:10.1098/rstb.2015.0360 (2016).

137. Mathis, A. *et al.* DeepLabCut: markerless pose estimation of user-defined body parts with deep learning. *Nature Neuroscience* **21,** 1281–1289. ISSN: 1097-6256, 1546-1726. doi:10.1038/s41593-018-0209-y (2018).

138. Lucas, B. D. & Kanade, T. *An Iterative Image Registration Technique with an Application to Stereo Vision* in *Proceedings of the 7th international joint conference on Artificial intelligence* Type: Conference Proceedings ().

139. Wang, Q. *et al.* The Allen Mouse Brain Common Coordinate Framework: A 3D Reference Atlas. *Cell* **181.** Publisher: Elsevier, 936–953.e20. ISSN: 0092-8674, 1097-4172. doi:10.1016/j.cell.2020.04.007 (2020).

140. Saxena, S. *et al.* Localized semi-nonnegative matrix factorization (LocaNMF) of widefield calcium imaging data. *PLOS Computational Biology* **16** (ed Macke, J. H.) e1007791. ISSN: 1553-7358. doi:10.1371/journal.pcbi.1007791 (2020).

141. Virtanen, P. *et al.* SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods* **17.** Number: 3 Publisher: Nature Publishing Group, 261–272. ISSN: 1548-7105. doi:10.1038/s41592-019-0686-2 (2020).

142. Denovellis, E. *Eden-Kramer-Lab/ripple_detection: v1.2.0* version v1.2.0. 2022. doi:10.5281/zenodo.7011175.

143. Kay, K. *et al.* A hippocampal network for spatial coding during immobility and sleep. *Nature* **531.** Number: 7593 Publisher: Nature Publishing Group, 185–190. ISSN: 1476-4687. doi:10.1038/nature17144 (2016).

144. Gibson, J. F., Doyne Farmer, J., Casdagli, M. & Eubank, S. An analytic approach to practical state space reconstruction. *Physica D: Nonlinear Phenomena* **57,** 1–30. ISSN: 01672789. doi:10.1016/0167-2789(92)90085-2 (1992).

145. Broomhead, D. & King, G. P. Extracting qualitative dynamics from experimental data. *Physica D: Nonlinear Phenomena* **20,** 217–236. ISSN: 01672789. doi:10.1016/0167-2789(86)90031-X (1986).

146. Kingma, D. P. & Welling, M. *Auto-Encoding Variational Bayes* 2014. arXiv: 1312.6114[cs,stat].

147. Sønderby, C. K., Raiko, T., Maaløe, L., Sønderby, S. K. & Winther, O. *Ladder Variational Autoencoders* 2016. doi:10.48550/arXiv.1602.02282. arXiv: 1602.02282[cs,stat].

148. Higgins, I. *et al. beta-VAE: Learning Basic Visual Concepts with a Constrained Variational Framework* in. International Conference on Learning Representations (2022).

149. Casdagli, M., Eubank, S., Farmer, J. D. & Gibson, J. State space reconstruction in the presence of noise. *Physica D: Nonlinear Phenomena* **51,** 52–98. ISSN: 0167-2789. doi:10.1016/0167-2789(91)90222-U (1991).

150. Kingma, D. P. & Ba, J. *Adam: A Method for Stochastic Optimization* 2017. arXiv: 1412.6980[cs].

151. Bowman, S. R. *et al. Generating Sentences from a Continuous Space* 2016. doi:10.48550/arXiv.1511.06349. arXiv: 1511.06349[cs].

152. Halko, N., Martinsson, P. G. & Tropp, J. A. Finding Structure with Randomness: Probabilistic Algorithms for Constructing Approximate Matrix Decompositions. *SIAM Review* **53.** Publisher: Society for Industrial and Applied Mathematics, 217–288. ISSN: 0036-1445. doi:10.1137/090771806 (2011).

153. Schmid, P. J. Dynamic mode decomposition of numerical and experimental data. *Journal of Fluid Mechanics* **656,** 5–28. ISSN: 0022-1120, 1469-7645. doi:10.1017/S0022112010001217 (2010).

154. Kutz, J. N., Brunton, S. L., Brunton, B. W. & Proctor, J. L. *Dynamic Mode Decomposition: Data-Driven Modeling of Complex Systems* Google-Books-ID: kL_zDQAAQBAJ. 241 pp. ISBN: 978-1-61197-449-2 (SIAM, 2016).

155. Colbrook, M. J. *The Multiverse of Dynamic Mode Decomposition Algorithms* 2023. doi:10.48550/arXiv.2312.00137. arXiv: 2312.00137.

156. Erichson, N. B., Mathelin, L., Kutz, J. N. & Brunton, S. L. Randomized Dynamic Mode Decomposition. *SIAM Journal on Applied Dynamical Systems* **18.** Publisher: Society for Industrial and Applied Mathematics, 1867–1891. doi:10.1137/18M1215013 (2019).

157. Demo, N., Tezzele, M. & Rozza, G. PyDMD: Python Dynamic Mode Decomposition. *Journal of Open Source Software* **3,** 530. ISSN: 2475-9066. doi:10.21105/joss.00530 (2018).

158. Ichinaga, S. M. *et al. PyDMD: A Python package for robust dynamic mode decomposition* 2024. doi:10.48550/arXiv.2402.07463. arXiv: 2402.07463.

159. Jammalamadaka, S. R. & Sarma, Y. R. *A correlation coefficient for angular variables* in *Statistical theory and data analysis II* Statistical theory and data analysis II (1988), 349–364.

160. Pedregosa, F. *et al.* Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* **12,** 2825–2830. ISSN: 1533-7928 (2011).

161. Serneels, S., De Nolf, E. & Van Espen, P. J. Spatial Sign Preprocessing: A Simple Way To Impart Moderate Robustness to Multivariate Estimators. *Journal of Chemical Information and Modeling* **46,** 1402–1409. ISSN: 1549-9596, 1549-960X. doi:10.1021/ci050498u (2006).

162. *hmmlearn* https://github.com/hmmlearn/hmmlearn. 2025.

163. . *mouse silhouette* 2020. doi:10.5281/zenodo.3925991.

164. Williams, M. *et al.* A kernel-based method for data-driven koopman spectral analysis. *Journal of Computational Dynamics* **2,** 247–265. ISSN: 2158-2505. doi:10.3934/jcd.2015005 (2015).

165. Lazari, A. *et al.* The mouse motor system contains multiple premotor areas and partially follows human organizational principles. *Cell Reports* **43.** Publisher: Elsevier. ISSN: 2211-1247. doi:10.1016/j.celrep.2024.114191 (2024).

Supplementary Text

Toy example: Stochastic Lorenz system

We demonstrate the efficacy of our procedure on the three-dimensional Lorenz system, a canonical low-dimensional nonlinear dynamical system. Specifically, we simulated the following stochastic system:

$$\begin{aligned} \mathrm{d}z_1 &= 10(z_2 - z_1)\mathrm{d}t + \alpha \mathrm{d}W_1, \\ \mathrm{d}z_2 &= (z_1(28 - z_3) - z_2)\mathrm{d}t + \alpha \mathrm{d}W_2, \\ \mathrm{d}z_3 &= (z_1 z_2 - \frac{8}{3} z_3)\mathrm{d}t + \alpha \mathrm{d}W_3, \end{aligned} \tag{16}$$

where coefficients reflect standard parameter choices, and α is the scaling parameter applied to three independent Wiener processes (W_1, W_2, W_3) acting on the three state variables.

For all simulations, the observation functions g_n are simply identity functions, the scalar observable y_1 (by analogy, pupil diameter) is a function of the state variable z_2, and the multidimensional observable $\mathbf{y}_2$ (by analogy, a brain image timecourse) is a function of the other two state variables, $[z_1, z_3]^{\mathrm{T}}$). We considered multiple levels of dynamical noise $\omega(t)$, which was modeled as a multidimensional Wiener process scaled by the parameter α, and multiple levels of the observation noises σ_1 and σ_2, which were modeled as unit Gaussians multiplied by a scaling factor β. Thus, the stochastic Lorenz governing equations (Eq. 16) were coupled to the observation functions:

$$y_1 = g_1(z_2) + \beta\sigma_1 = z_2 + \beta\sigma_1, \tag{17}$$

$$\mathbf{y}_2 = \mathbf{g}_2\left(\begin{bmatrix} z_1 \\ z_3 \end{bmatrix}\right) + \beta\sigma_2 = \begin{bmatrix} z_1 \\ z_3 \end{bmatrix} + \beta\sigma_2. \tag{18}$$

We sought to reconstruct the dynamics of the 2D observable $\mathbf{y}_2$ on the basis of a time delay embedding of the scalar observable y_1. As in the main text analyses, we assessed reconstruction accuracy in test data as a function of whether delay coordinates and/or nonlinearities were included in the model.

Lorenz trajectories were simulated with $\Delta t = .01$ and $N = 10000$ time steps. The first 6000 time steps were included in the training set, and the model was tested on the final 3000 time steps. Stochastic trajectories (i.e., where $\alpha \neq 0$) were simulated with the standard Euler-Maruyama integration scheme.

Fig. S3 demonstrates the efficacy of this procedure across a range of values for the dynamical noise and observation noise (controlled by parameters α and β, respectively). With no dynamical noise or observation noise, the "Latent model" enabled nearly perfect prediction of the multidimensional observable on the basis of the scalar observable y_1 and its dynamics. Delay embedding and nonlinearities were generally both crucial to optimal performance. Across all models, performance gradually degraded with increasing levels of dynamical noise (Fig. S3B-C). However, notably, performance could remain high even in the presence of dynamical noise that introduces obvious distortion to the latent trajectories. This is because the noise is broadcast throughout the system, and is thus manifested in the dynamics of both observables.

On the other hand, with increasing levels of observation noise (increasing β), which is derived from Gaussian fluctuations unique to y_1 and $\mathbf{y}_2$, the proportion of predictable variance in $\mathbf{y}_2$ rapidly decreased (Fig. S3 D-E). This is because variance in $\mathbf{y}_2$ became increasingly dominated by fluctuations that are not manifested in y_1, precluding the ability to model $\mathbf{y}_2$ via a time-invariant mapping from a state-space reconstruction based on y_1.

In the context of arousal dynamics, observation noise in our framework includes genuine measurement noise (e.g., imaging artifacts) as well as physiological sources of variance that are conditionally independent of arousal (e.g., fluctuations resulting from sensory-evoked neural activity, or processes other than "arousal" that are hypothetically reflected in pupil diameter [130]). Accordingly, if model performance is poor, this could reflect either genuine measurement noise or the predominance of physiological processes that are not common to both observables. By contrast, the strong performance of our method is an indication of the relatively minor contribution from both of these sources of observation noise in comparison to a dominant, shared latent process.

Universal embeddings via Legendre polynomials

The leading left eigenvectors of $\mathbf{H}$ approximate eigenfunctions of the Koopman operator $\mathcal{K}$, which advances measurements of a dynamical system [70]. Importantly, the eigenfunctions of $\mathcal{K}$ define a subspace invariant to the dynamics. These eigenfunctions can be used to parameterize the manifold based on linearization around a fixed point, extended throughout the basin of attraction [164]. Thus, whereas Results in Fig. 4 reflect mappings from a 2D latent space obtained through a neural network, these properties of Koopman eigenfunctions imply the possibility of using the leading Legendre coordinates (Eq. (5)) as intrinsic, analytical coordinates for parameterizing the arousal manifold $\mathcal{M}$.

Supplementary Figures

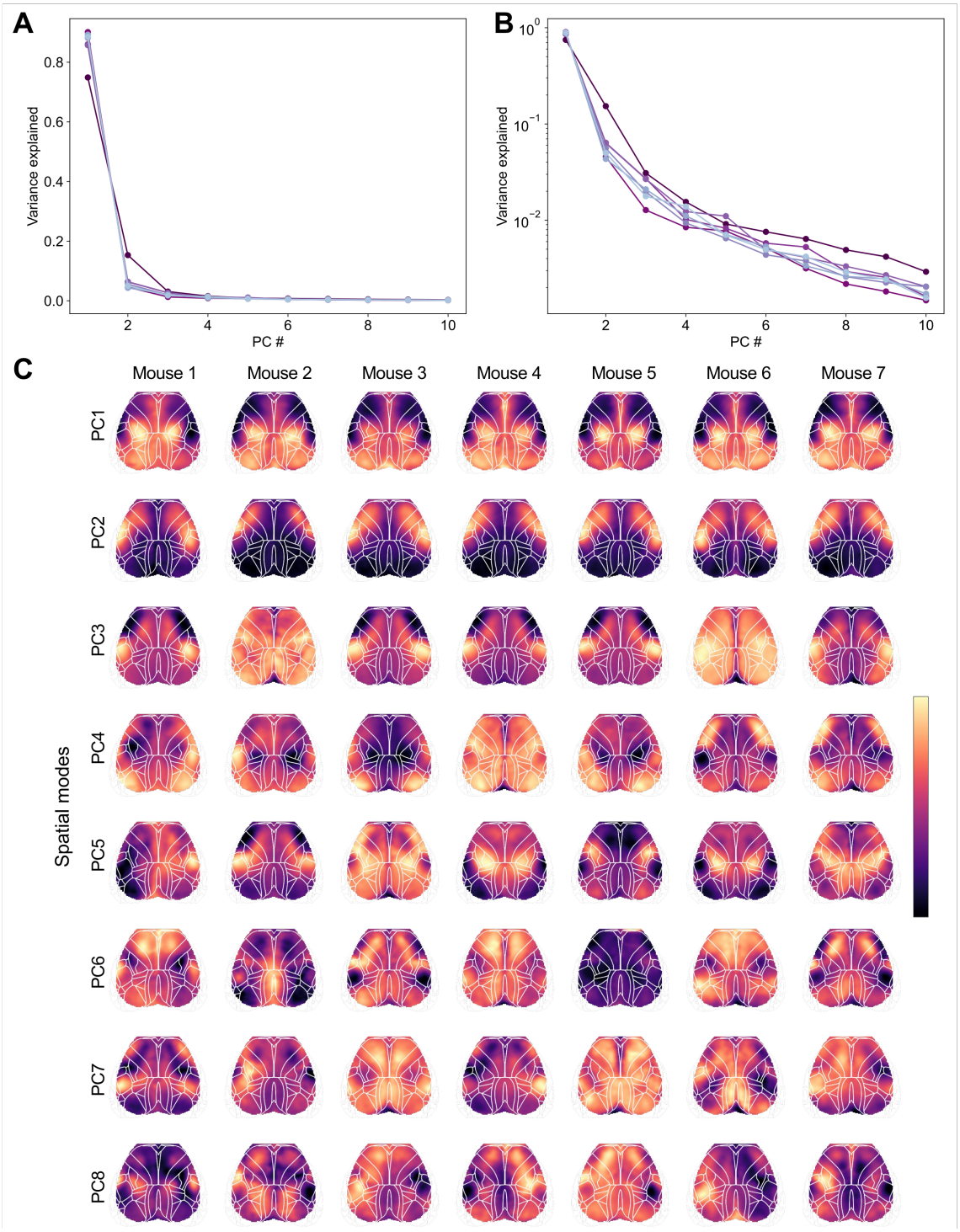

Figure S1: SVD results for widefield calcium images. **A-B** Scree plots for each of the seven mice, displayed on a linear (A) or log-transformed (B) y-axis. For all mice, images were dominated by the first principal component. **C** Spatial topographies corresponding to the top 8 PCs (arbitrary units). As the sign of each PC is arbitrary, some PCs were sign-flipped to improve visual correspondence across mice.

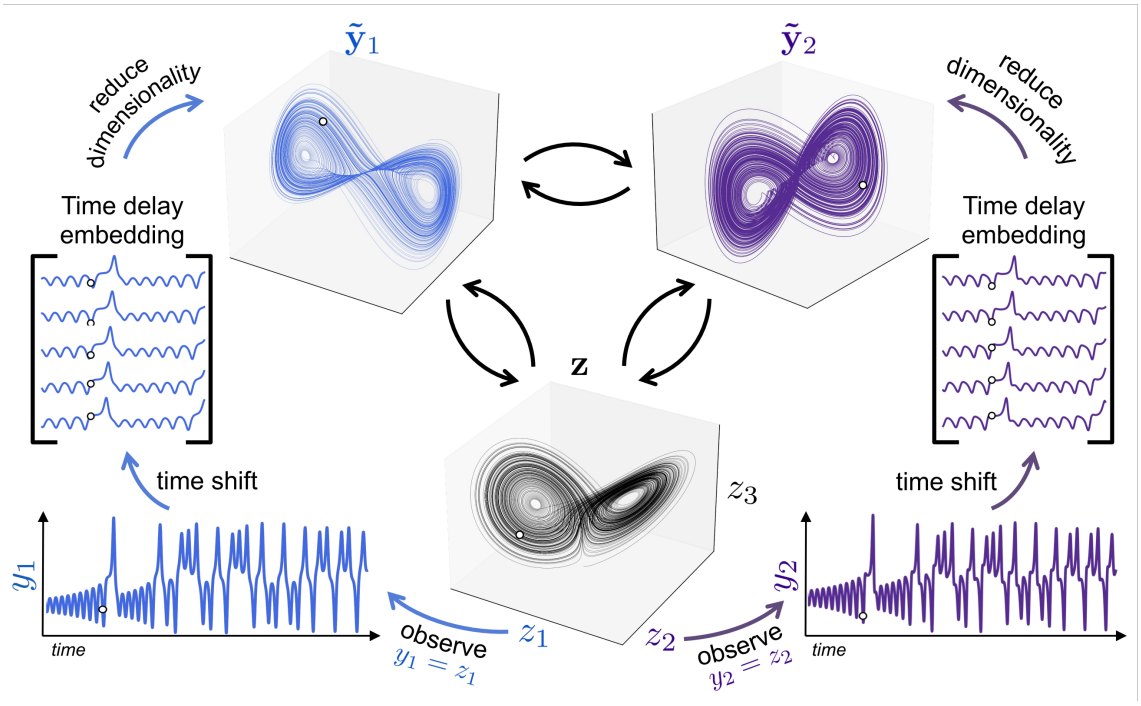

Figure S2: Illustration of shadow manifolds in the context of the 3D Lorenz system. The white dot in each plot represents the same point in time (i.e., the same dynamical state) from several different views (or embeddings). Here, the measurements y_1 and y_2 are simply taken to be the state variables z_1 and z_2 from the original system $\mathbf{z}(t)$. Each of these observations can be stacked with time-shifted copies of itself to obtain a (high-dimensional) time delay embedding that is diffeomorphic to the original attractor. These embeddings can be dimensionally reduced to obtain low-dimensional "shadow manifolds" that (approximately) preserve the topology of the original system. Pairs of black arrows indicate relation via a diffeomorphic mapping. These relations clarify the motivation for modeling one observable from a dynamical system (e.g., y_2) as a function of a time delay embedding of a second observable (e.g., the augmented observable $\tilde{\mathbf{y}}_1$). See Fig. S3 for demonstration of our computational framework applied to the Lorenz system.

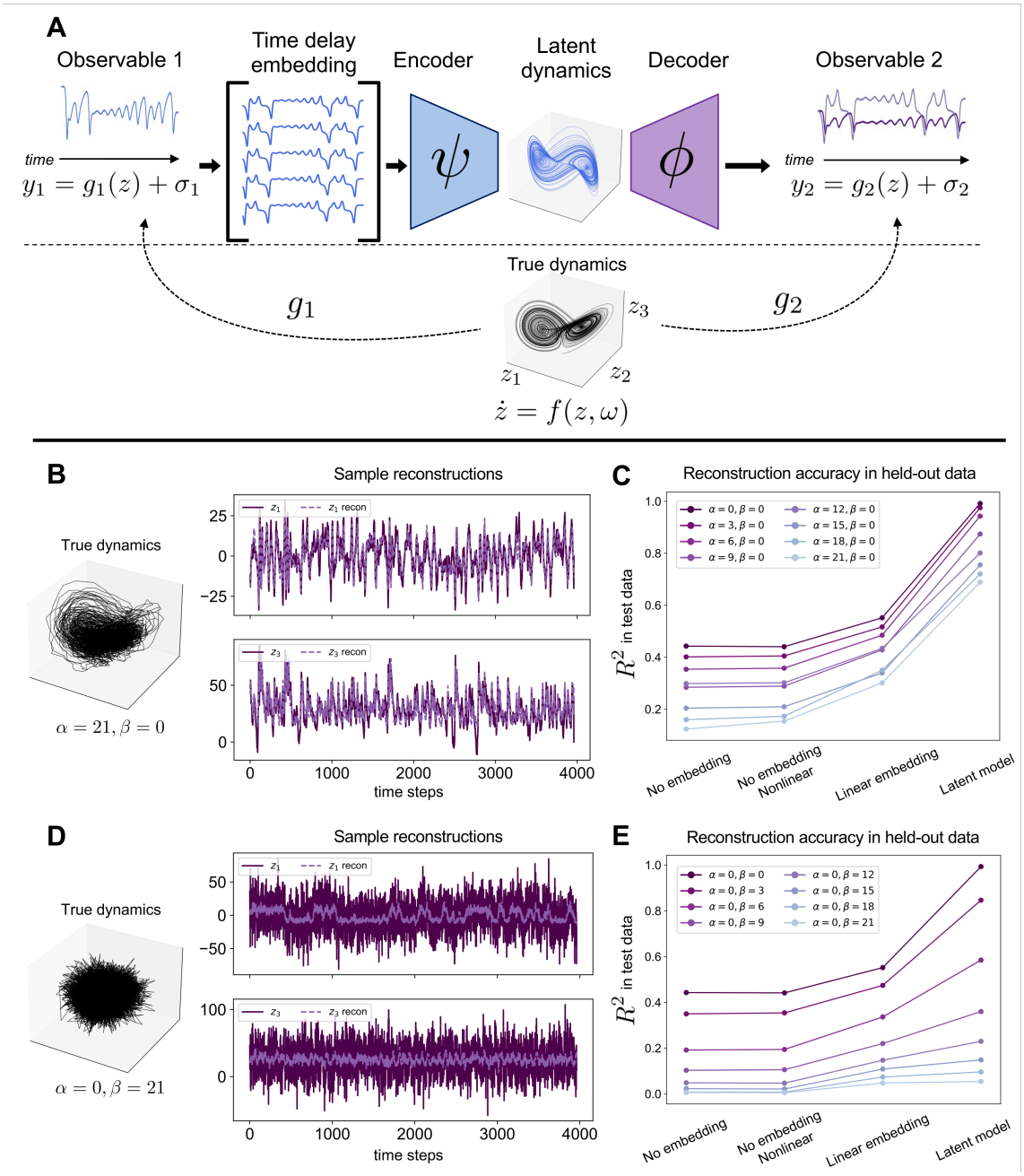

Figure S3: Toy demonstration of delay embedding for multimodal cross-prediction in the presence of dynamical and observation noise. **A** Delay embedding framework applied to a stochastic version of the 3D Lorenz system (compare to Fig. 1C). Here, the observation functions g_1 and g_2 are defined as identity functions. Across simulations, y_1 is taken to be a function of z_2, whereas $\mathbf{y}_2$ is a 2D function of z_1 and z_3. We aim to model $\mathbf{y}_2$ according to y_1 (that is, a noisy measurement of a single state variable, z_2), across different levels of dynamical noise and observation noise. f corresponds to the stochastic Lorenz system defined in Eq. 16. **B-C** Reconstruction accuracy across multiple levels of dynamical noise, scaled by the parameter α. B illustrates sample reconstructions of the state variables z_1 and z_3 from the observable y_1 at the highest dynamical noise level tested ($\alpha = 21$). **D-E** Same as B-C, but across multiple levels of observation noise, scaled by the parameter β. D illustrates sample reconstructions for the state variable time series at the highest noise level of observation noise tested ($\beta = 21$). See Supplementary Text for further details and interpretation.

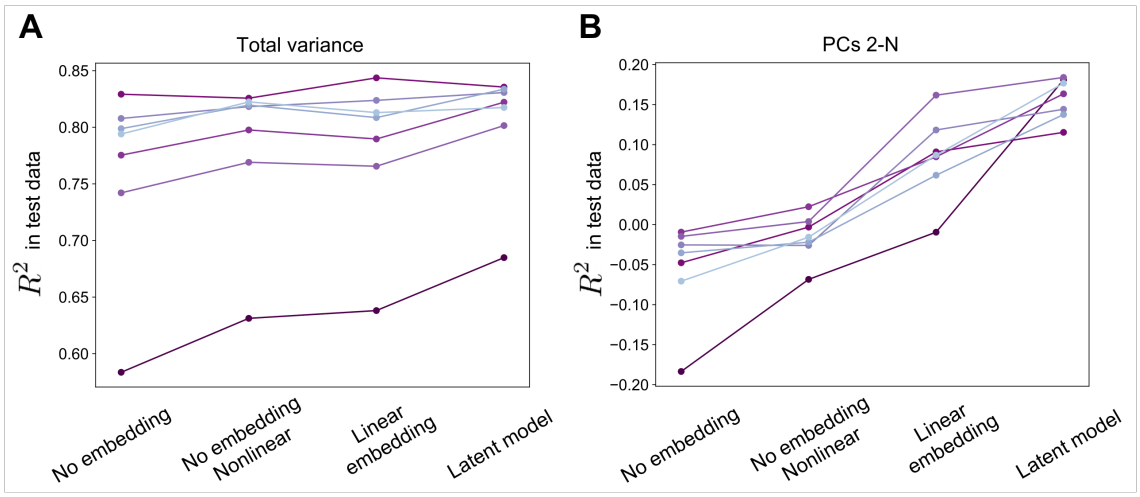

Figure S4: Variance explained in test data for four model categories: "No embedding" (linear pupil regressor), "No embedding, Nonlinear" (nonlinear mapping from pupil regressor), "Linear embedding" (linear mapping from delay embedded pupil), and "Latent model" (nonlinear mapping from delay embedded pupil). Left, total variance explained; right, variance explained along PCs 2-N. Compare also with simulation in Fig. S3 C&E.

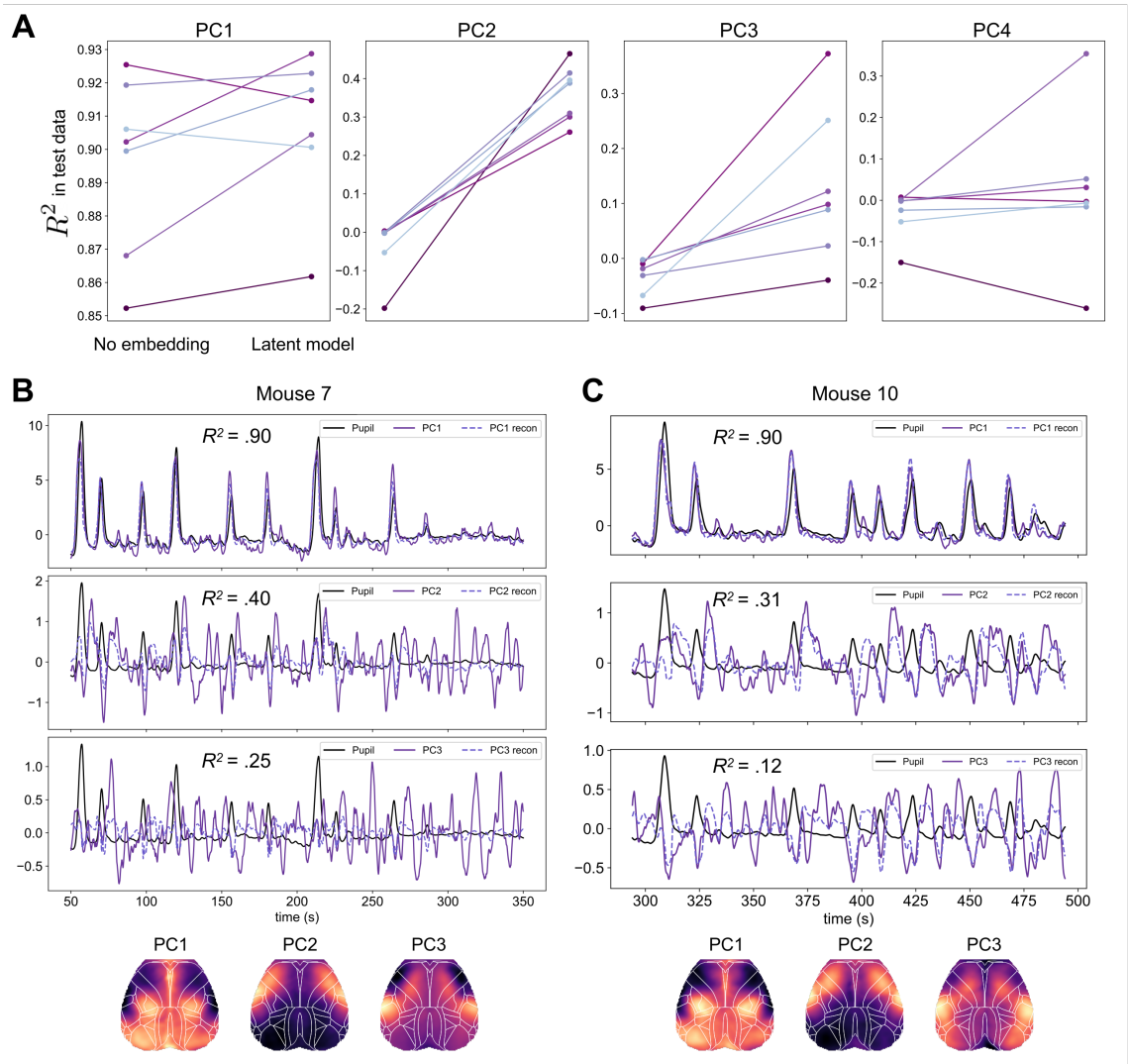

Figure S5: Pupil delay embeddings explain multidimensional variance in widefield calcium images. **A** Variance explained along the dimensions spanned by each of the first four principal components from each mouse. **B-C** Principal component (PC) time series and their prediction from pupil delay embeddings in test data, shown for two example mice. PC spatial maps corresponding to each time series are shown underneath. Higher-order PCs (here, 2 and 3) retain clear temporal relationships to PC1, enabling some of this multidimensional variance to be captured from a time-invariant mapping from the pupil-derived latent space.

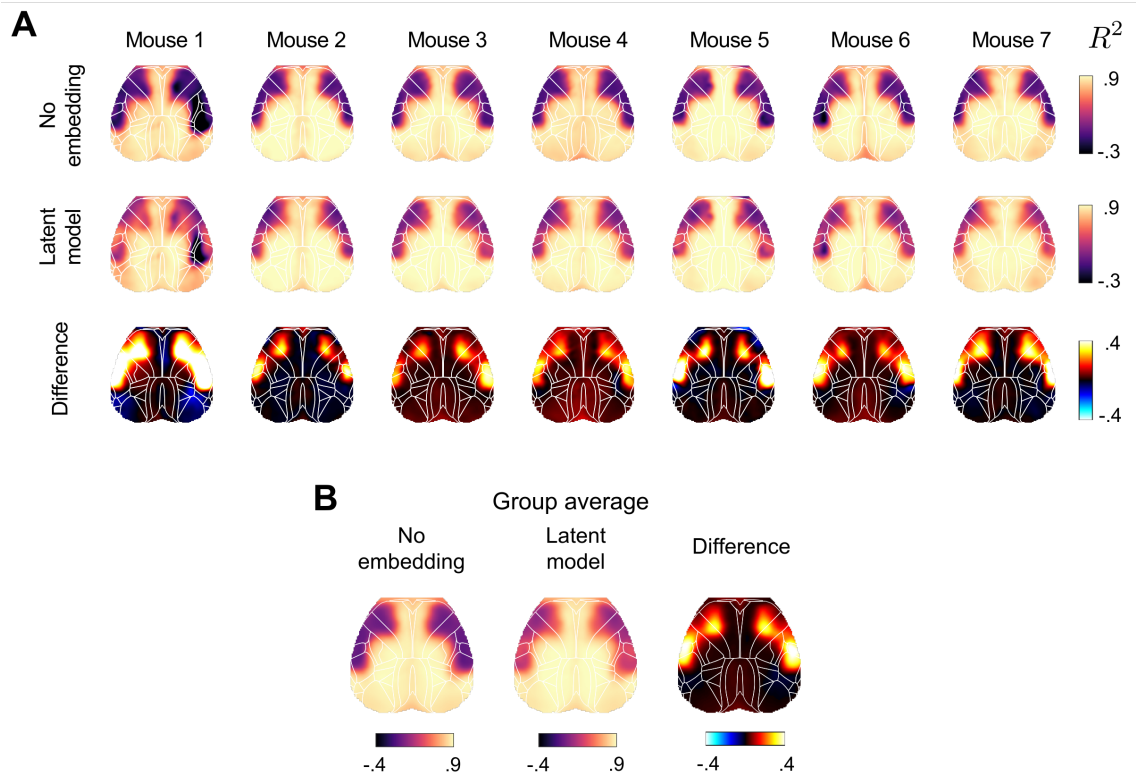

Figure S6: R^2 maps in held-out data corresponding to the "No embedding" and "Latent model" predictions, shown for each mouse (A) and averaged across mice (B). Dynamical information consistently improved explanatory power, particularly within a bilateral set of premotor cortical regions. This pattern recalls prior results in humans (cf. [117, 165]).

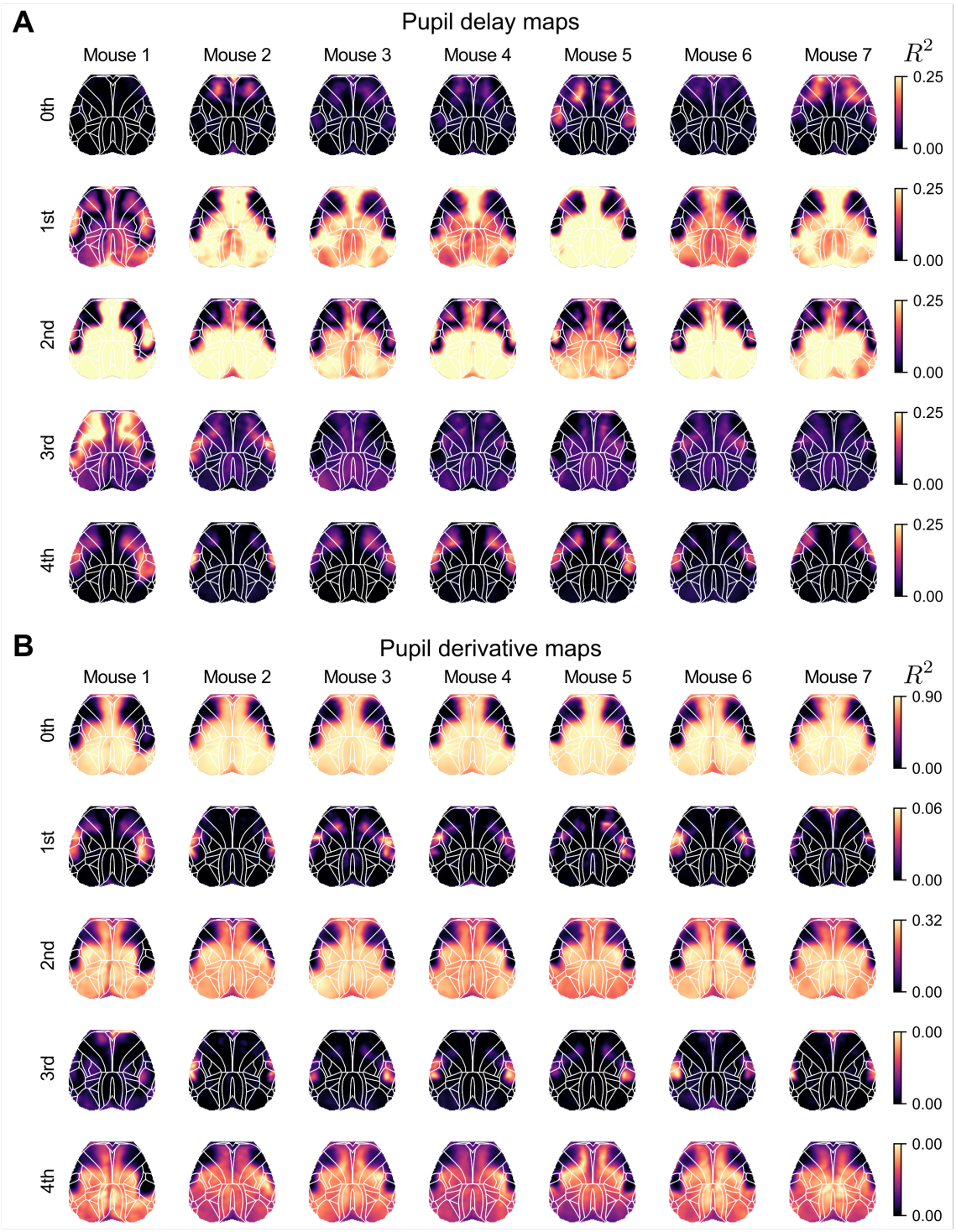

Figure S7: R^2 maps for widefield calcium images as linearly modeled by dynamical features of pupil size (no cross-validation). **A** R^2 maps derived from the pupil delay embedding (i.e., the Hankel matrix) projected onto successive Legendre polynomials. Projections onto Legendre polynomials (i.e., the Legendre coordinates $\tilde{\mathbf{y}}$ from Eq. (5)) essentially define convolutions of pupil diameter with increasingly high-frequency information from the recent past. The 0th-order polynomial effectively captures the amplitude envelope of pupil diameter (cf. widefield maps obtained from cholinergic indicators [20], which also correlate with pupil amplitude envelope [94]). **B** R^2 maps derived from successive temporal derivatives of pupil diameter (0th order derivative corresponds to the original pupil diameter time series).

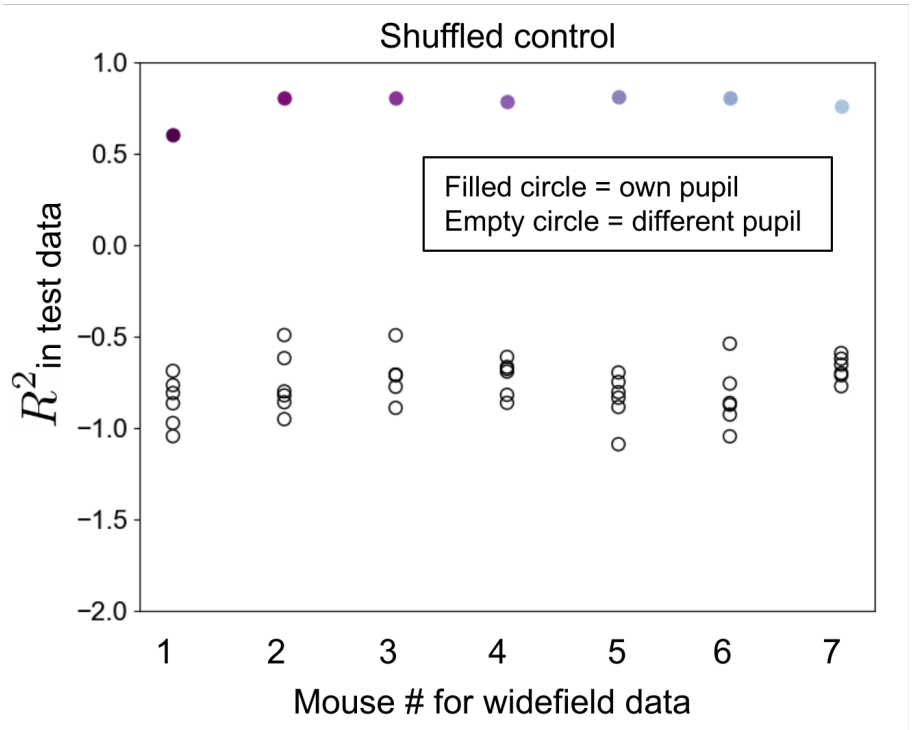

Figure S8: Shuffled control analysis for prediction of widefield calcium data from pupil dynamics (compare with Fig. 2A). Models were trained according to the standard leave-one-out pipeline. However, at test time, widefield data from the held-out mouse was predicted from not only the held-out mouse's own pupil data (filled circles), but also, separately, from the pupil data from each of the other mice (empty circles). For all mice, explained variance is high when using the mouse's own pupil timecourse (filled circles). In contrast, explained variance is invariably poor when attempting to predict widefield data from pupil dynamics obtained from another mouse (empty circles; negative R^2 values occur when the prediction performs worse than simply predicting the mean value of each pixel). This result clarifies that good performance in held-out data is not trivially obtained with any complex model.

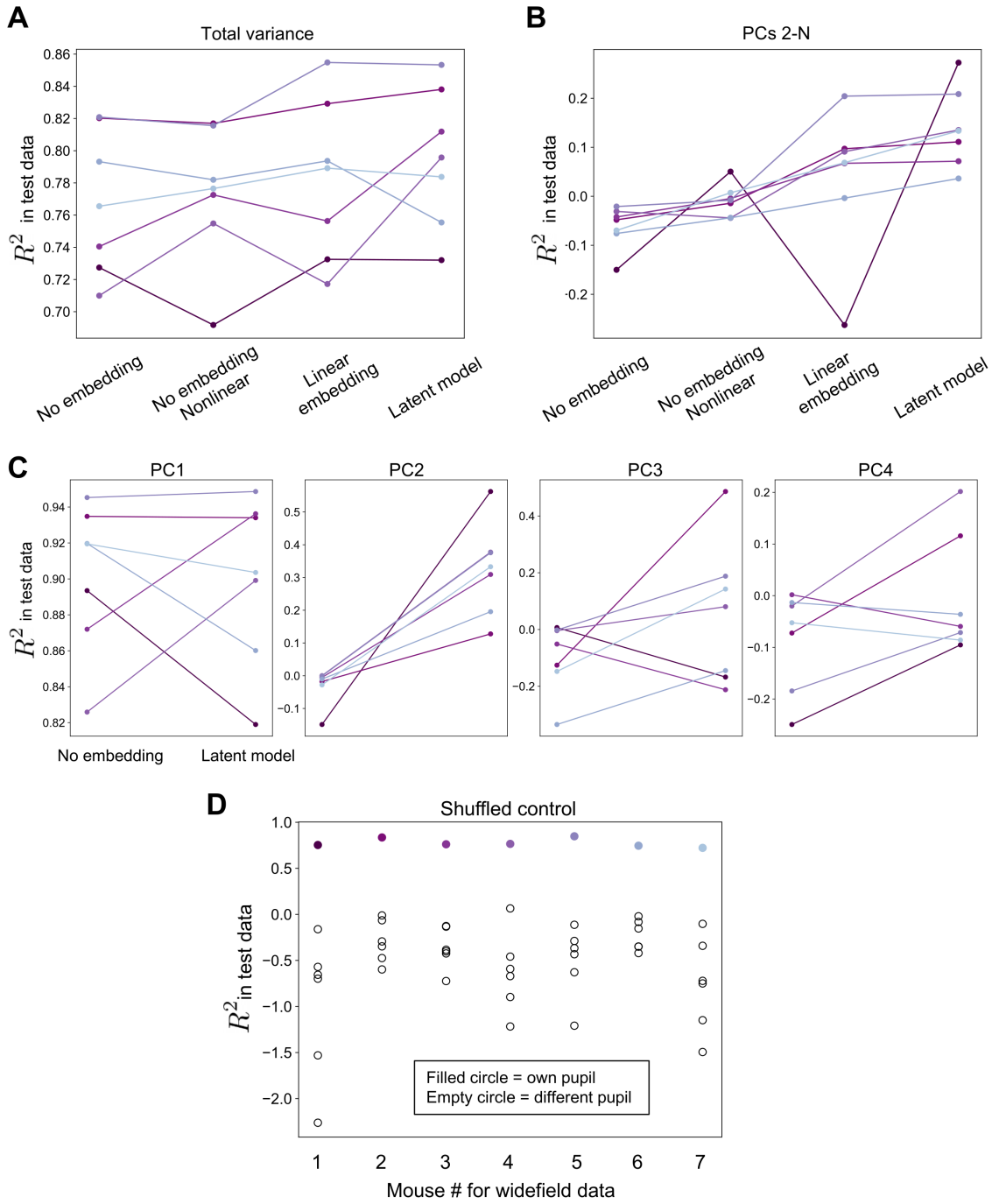

Figure S9: Results from the within-mouse model pipeline. Each mouse's dataset was split into train and test sets, as detailed in the Methods. **A-B** Same as in Fig. S4. **C** Same as in Fig. S5A. **D** Shuffled control analysis, as in Fig. S6, but adapted for the within-mouse model pipeline. Accordingly, for each mouse's test data, we trained seven models to predict the widefield data from pupil dynamics, with each model using the pupil time series from the training set of a different mouse. Explained variance (R^2) is reported for each model in test data. Filled circles correspond to models trained on the test mouse's own training data, as in A-C.

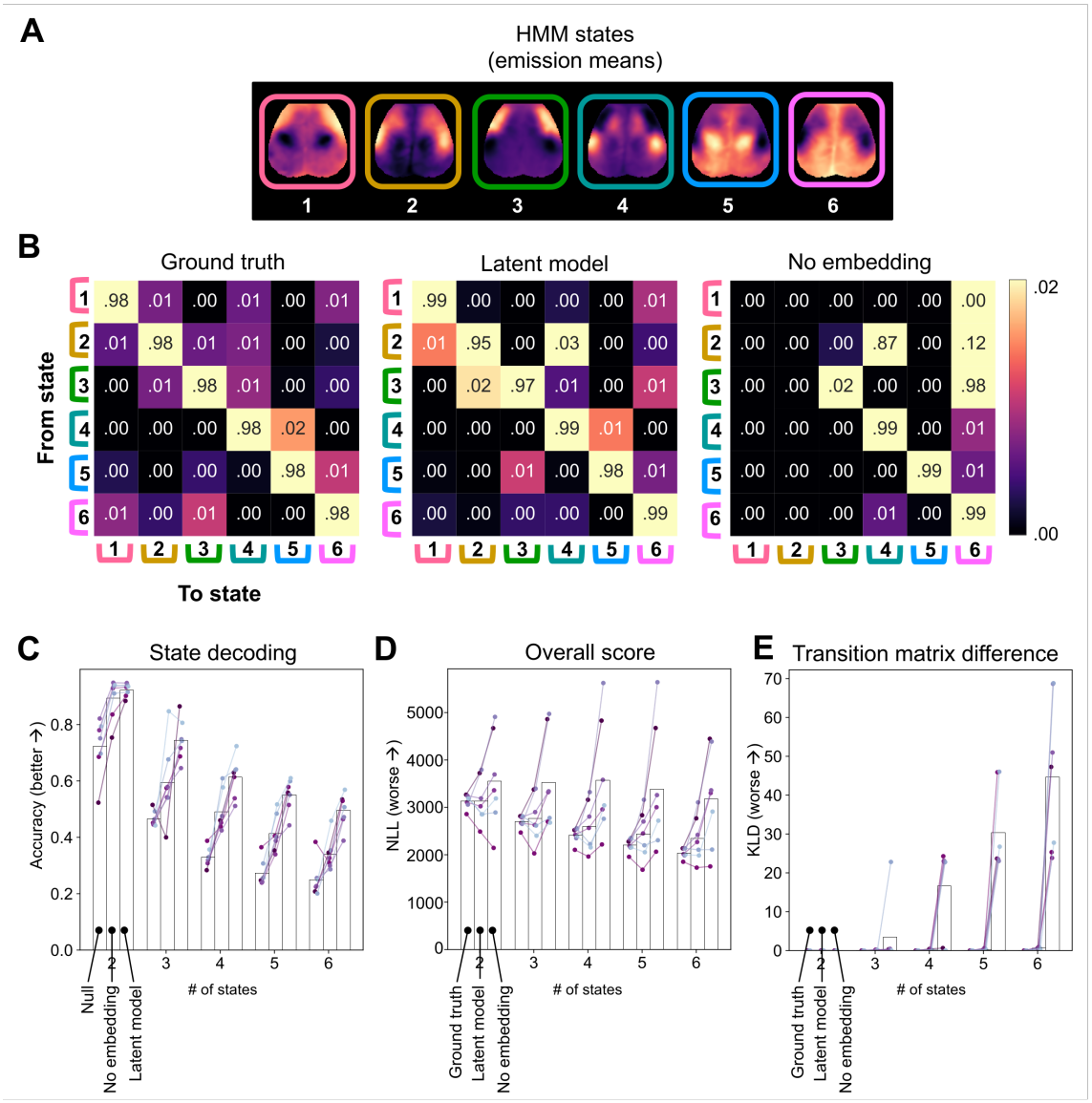

Figure S10: Hidden Markov models (HMMs) reveal that the latent model recapitulates spatially structured dynamics that are not captured by instantaneous pupil diameter. **A** Six HMM "states" for a sample mouse (same as in Fig. 2D), computed as the mean of the Gaussian emission associated with each hidden state. **B** HMM transition matrices involving the HMM states in A for the sample mouse, computed from the original data (left), the data reconstructed through the Latent model (middle), and the data reconstructed through the No embedding model (right). See Methods for model fitting details. Note high probabilities along the superdiagonal (elements immediately to the right of the principal diagonal) indicating a canonical cycle among states in this mouse (cf. Fig. 2D). This structure is less apparent in the No embedding reconstruction (right). **C-E** Evaluation metrics computed over a range of latent states (2-6) for all seven mice. **C** Decoding accuracy, as in Fig. 2E. Null predictions are based on repeatedly choosing the most common state in the ground truth HMM for each time point. Greater accuracy indicates greater agreement of latent state labels with labels obtained from the original HMM. **D** Total negative log-likelihood (NLL) of the ground truth or reconstructed datasets based on the HMM fit to the original data. Greater NLL indicates data are less probable according to ground truth generative model. **E** KL divergence of the ground truth and reconstructed transition matrices from the original HMM transition matrix. Greater KL divergence indicates greater difference from the ground truth transition matrix.

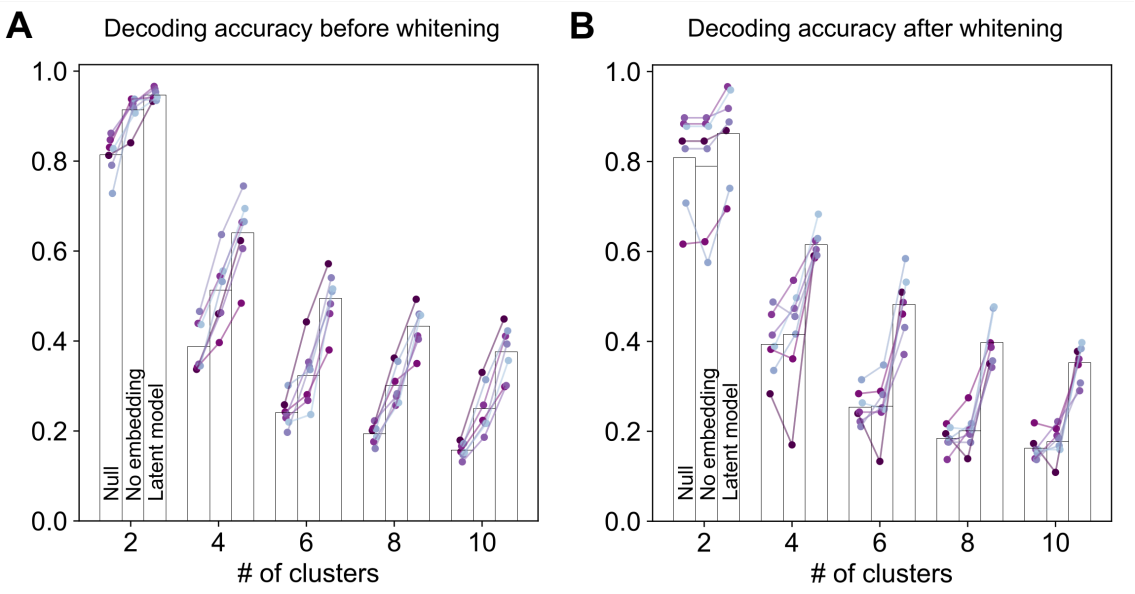

Figure S11: Decoding "brain states" from pupil dynamics. Comparison of decoder accuracy in properly identifying GMM cluster assignments for each image frame based on the most frequent assignment in the training set ("Null" decoder), the "No embedding" model, or the "Latent model". **A** Same as in Fig. 2E, but for up to 10 clusters and using a spherical rather than full covariance prior (Methods). **B** Same as in (A) but after whitening transformation (i.e., after projection onto the top 3 PCs, each dimension was set to unit variance). The whitening transformation highlights the ability of the delay embedding to capture genuinely multidimensional information, which is otherwise masked by the dominant PC1.

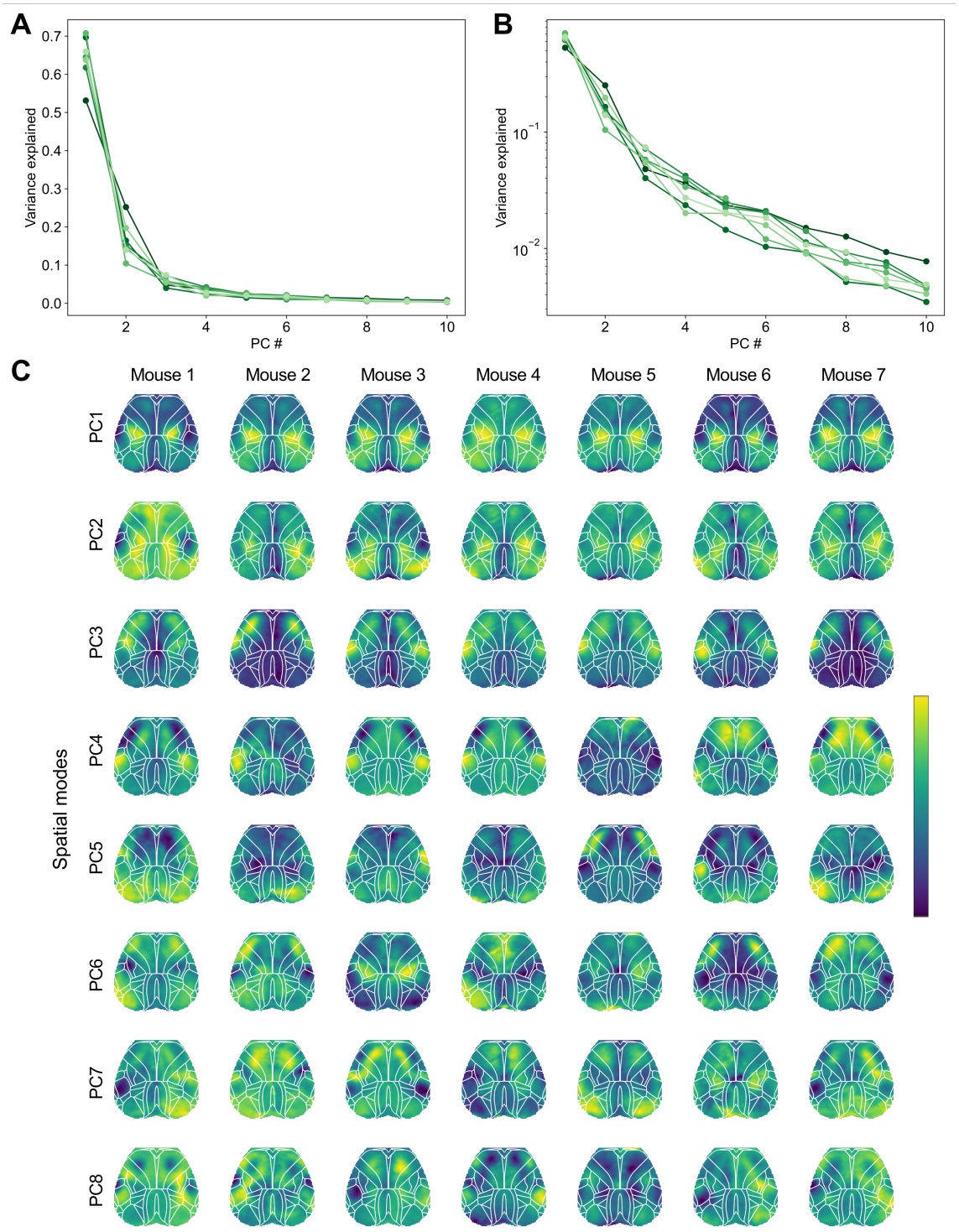

Figure S12: SVD results for widefield FAF images (as in Fig. S1). **A-B** Scree plots for each of the seven mice, displayed on a linear (A) or log-transformed (B) y-axis. For all mice, images were dominated by the first principal component. **C** Spatial topographies corresponding to the top 8 PCs (arbitrary units). As the sign of each PC is arbitrary, some PCs were sign-flipped to improve visual correspondence across mice.

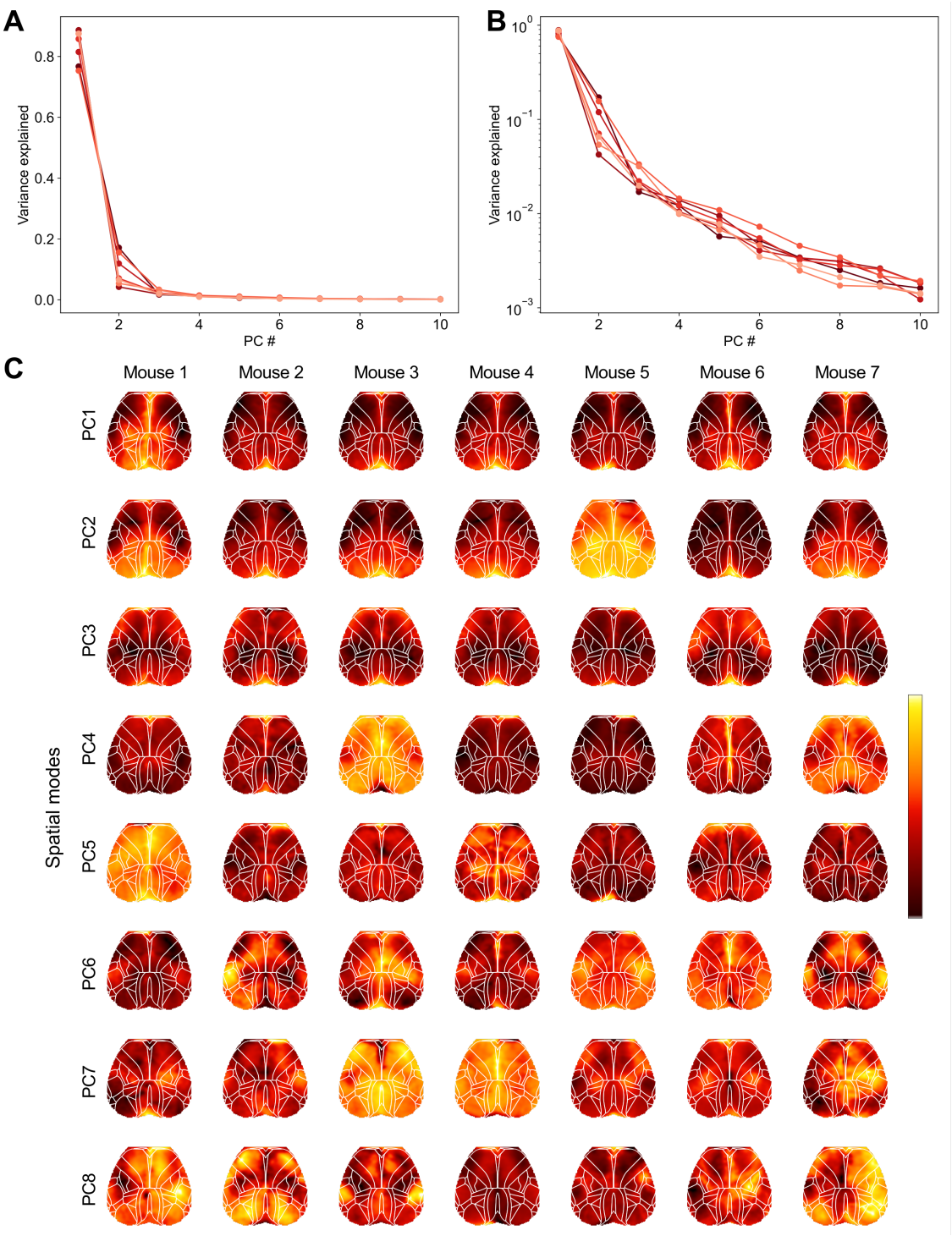

Figure S13: SVD results for widefield hemoglobin images (as in Figs. S1 & S12). **A-B** Scree plots for each of the seven mice, displayed on a linear (A) or log-transformed (B) y-axis. For all mice, images were dominated by the first principal component. **C** Spatial topographies corresponding to the top 8 PCs (arbitrary units). As the sign of each PC is arbitrary, some PCs were sign-flipped to improve visual correspondence across mice.

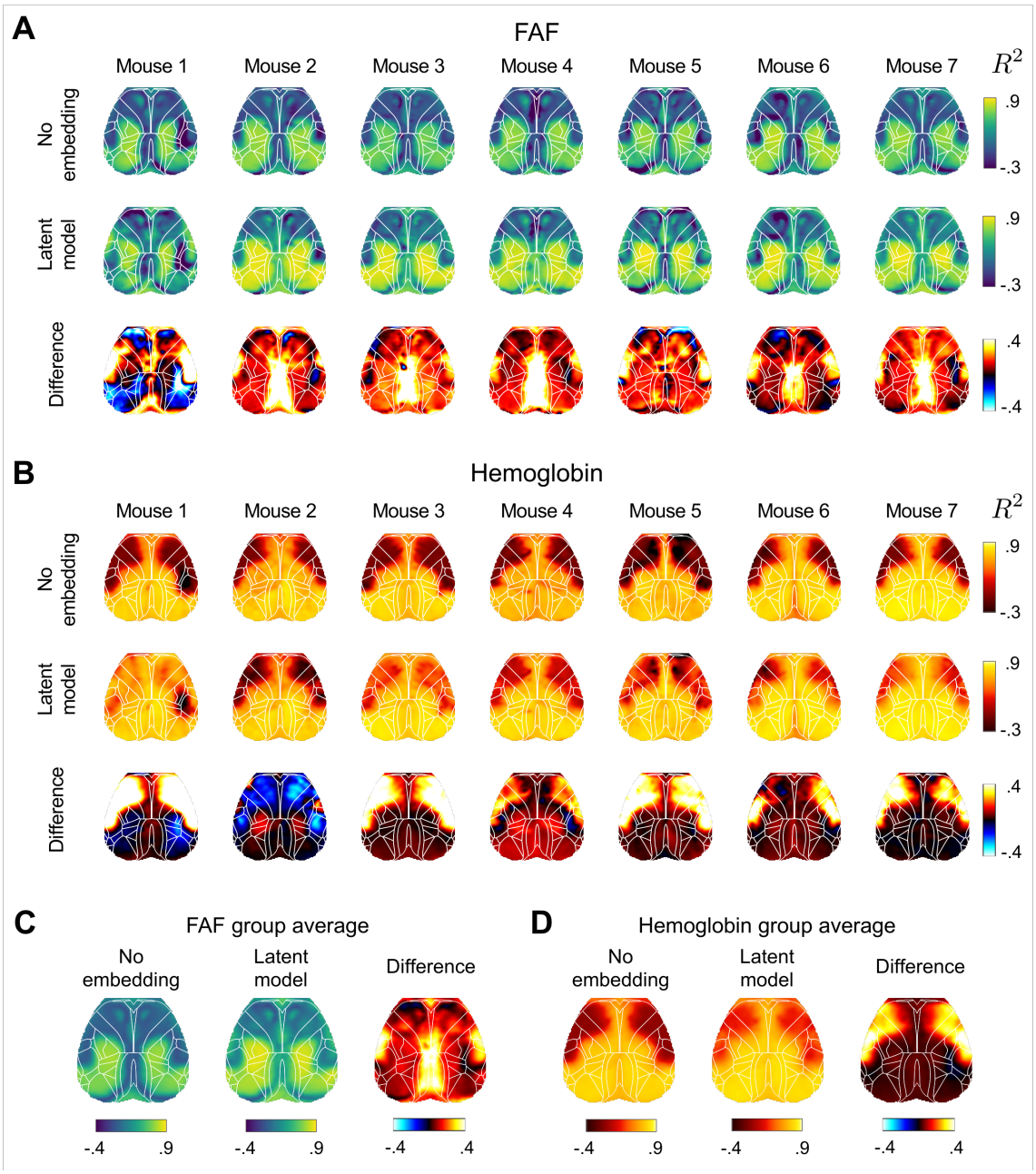

Figure S14: R^2 maps in held-out data corresponding to the "No embedding" and "Latent model" predictions, shown for FAF (A, C) and hemoglobin (B, D), for each mouse (A, B) and averaged across mice (C, D). See Fig. S6 for calcium results.

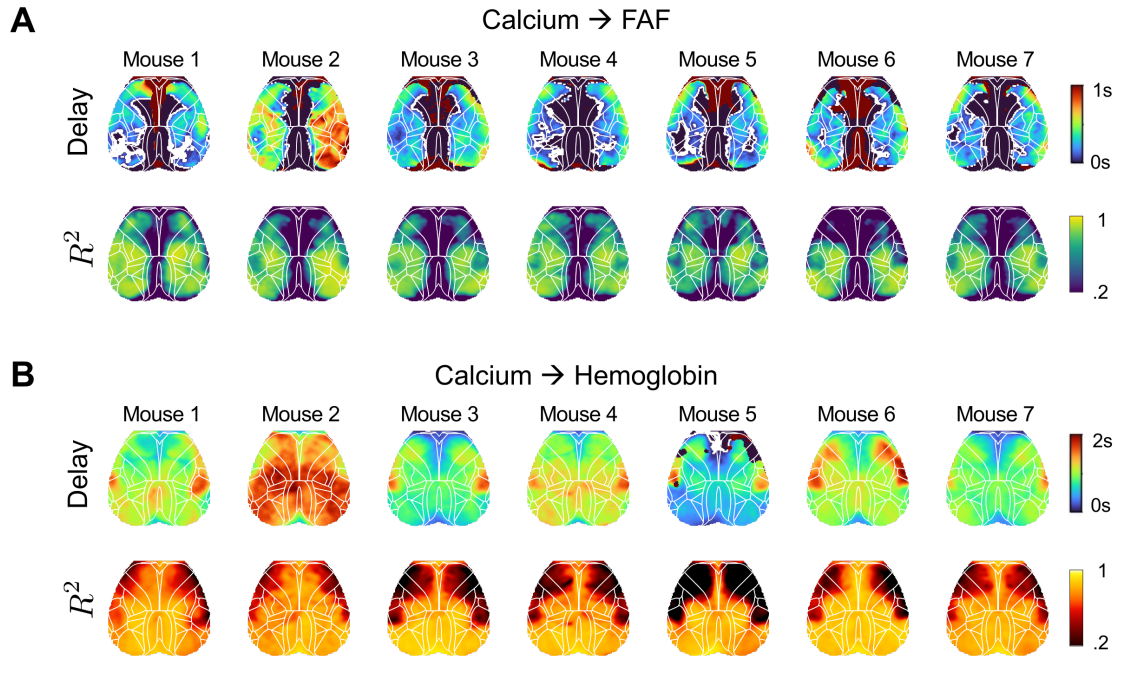

Figure S15: R^2 maps for widefield FAF (A) and hemoglobin (B) recordings linearly modeled at each pixel according to the widefield calcium measurements recorded at the same pixel. R^2 values (lower panels) are reported at the time shift (upper panels) optimizing the cross-correlation function between the calcium and FAF (or hemodynamic) recording at that pixel. Note that these model results are reported as is, i.e. without cross-validation. Extremal or missing values in the delay maps are generally due to the absence of a single, well-defined peak in the cross-correlation function within the allowable range (± 5 sec). This is commonly caused by biphasic patterns in the cross-correlation function. Note that delay embedding avoids this issue by combining multiple lagged copies of the predictor variable into a single, multidimensional predictor variable, rather than requiring a single lag-optimized predictor variable.

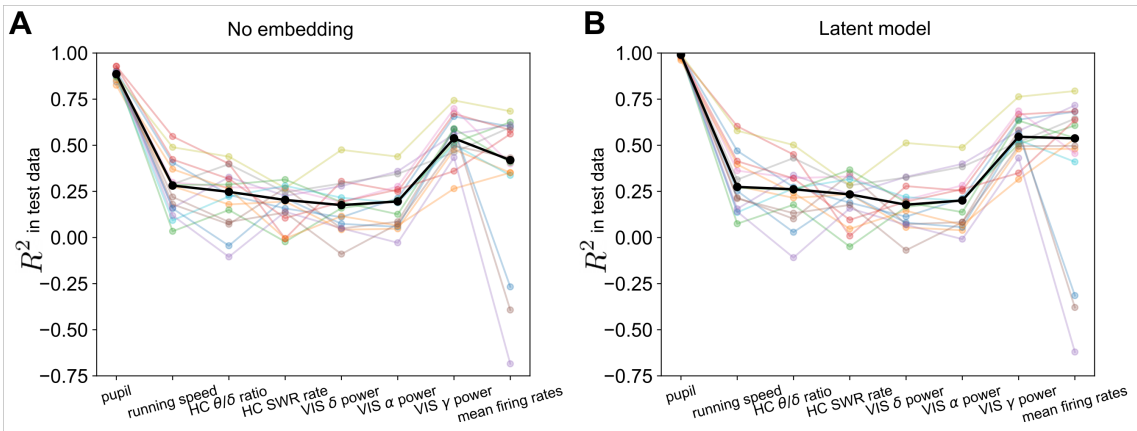

Figure S16: Leave-one-out cross validation analysis in Allen Brain Observatory data. Colored lines correspond to individual mice; black lines reflect median values across mice.