Supplementary Appendix

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:

$$\begin{aligned} \mathrm{d}z_1 &= 10(z_2 - z_1)\mathrm{d}t + \alpha \mathrm{d}W_1, \\ \mathrm{d}z_2 &= (z_1(28 - z_3) - z_2)\mathrm{d}t + \alpha \mathrm{d}W_2, \\ \mathrm{d}z_3 &= (z_1 z_2 - \frac{8}{3} z_3)\mathrm{d}t + \alpha \mathrm{d}W_3, \end{aligned} \tag{1}$$

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 (1) were coupled to the observation functions:

$$y_1 = g_1(z_2) + \beta\sigma_1 = z_2 + \beta\sigma_1, \tag{2}$$

$$\mathbf{y}_2 = g_2([z_1, z_3]^{\mathrm{T}}) + \beta\sigma_2 = [z_1, z_3]^{\mathrm{T}} + \beta\sigma_2. \tag{3}$$

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 derived from Gaussian fluctuations unique to y_1 and $\mathbf{y}_2$, the proportion of predictable variance in $\mathbf{y}_2$ rapidly decreased (Fig. S3D-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 [1]). 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 **H** approximate eigenfunctions of the Koopman operator $\mathcal{K}$, which advances measurements of a dynamical system [2]. 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 [3]. Thus, while 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 (**??**) as intrinsic, analytical coordinates for parameterizing the arousal manifold $\mathcal{M}$.

Supplementary Figures

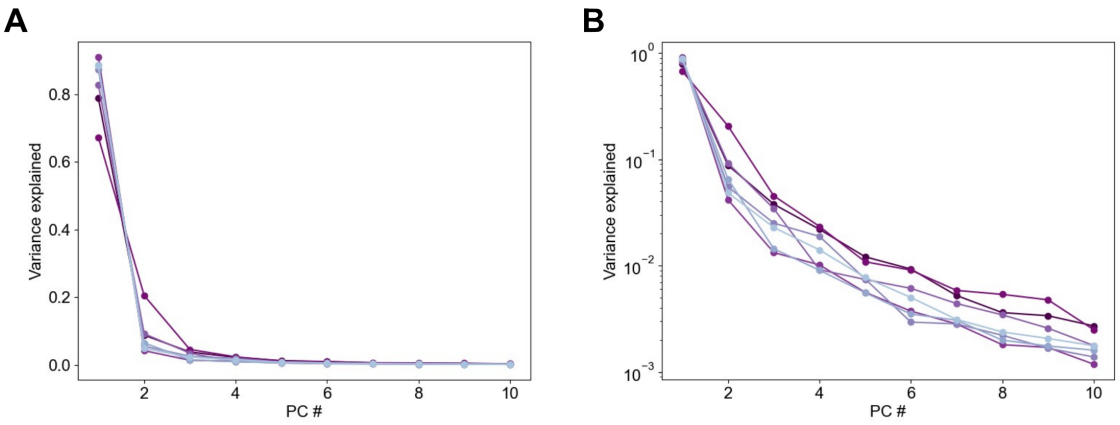

Figure S1: Scree plots for widefield calcium images from each of seven mice (right plot shows log-transformed y axis). For all mice, images were dominated by the first principal component.

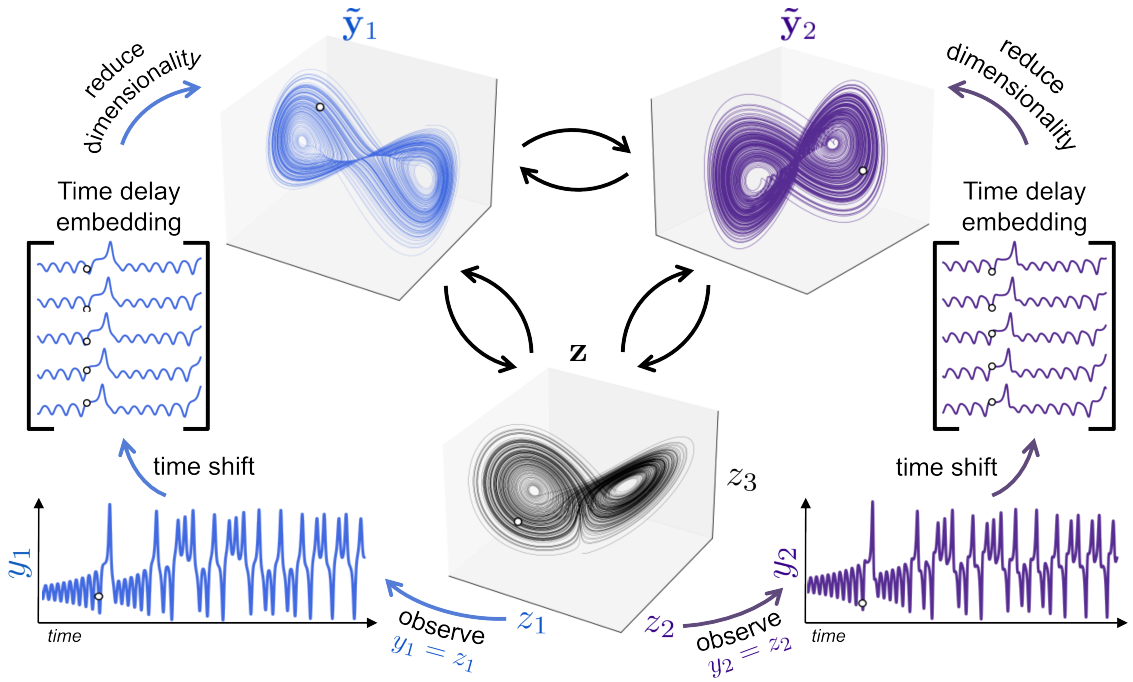

Figure S2: Illustration of shadow manifolds in the context of the 3D Lorenz system. The white dot in each plot represents the same point in time (i.e., the same dynamical state) from several different views (or embeddings). Here, the measurements y_1 and y_2 are simply taken to be the state variables 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.

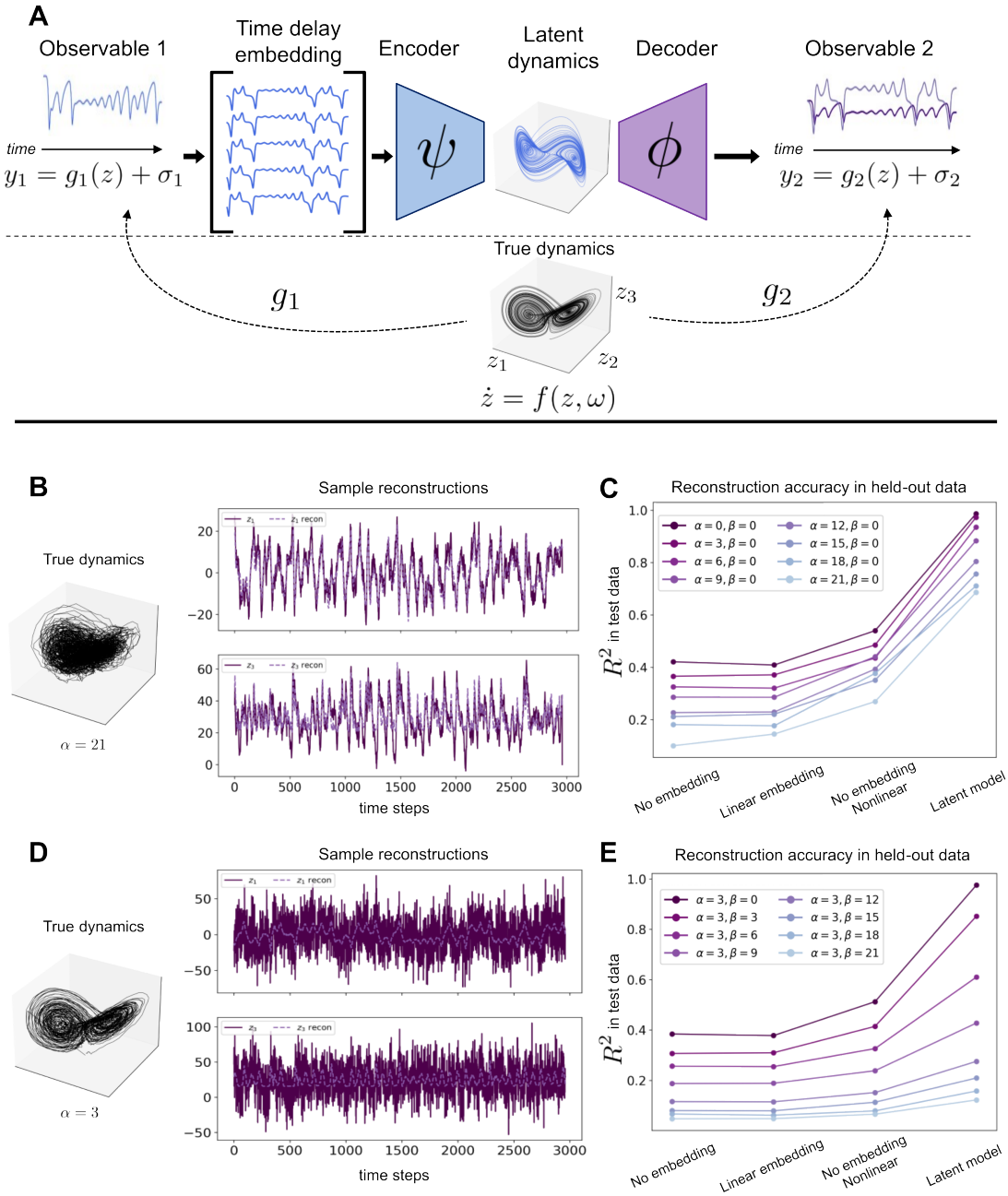

Figure S3: Toy demonstration of delay embedding for multimodal cross-prediction in the presence of dynamical and observation noise. **A** Delay embedding framework applied to a stochastic version of the 3D Lorenz system (compare to Fig. 1C). Here, the observation functions g_1 and g_2 are defined as identity functions. Across simulations, y_1 is taken to be a function of z_2, while $\mathbf{y}_2$ 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 one state variable, z_2), across different levels of dynamical noise and observation noise. f corresponds to the stochastic Lorenz system defined in 1. **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 , with a fixed level of dynamical noise ($\alpha = 3$). D illustrates latent dynamics for this level of dynamical noise, as well as 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.

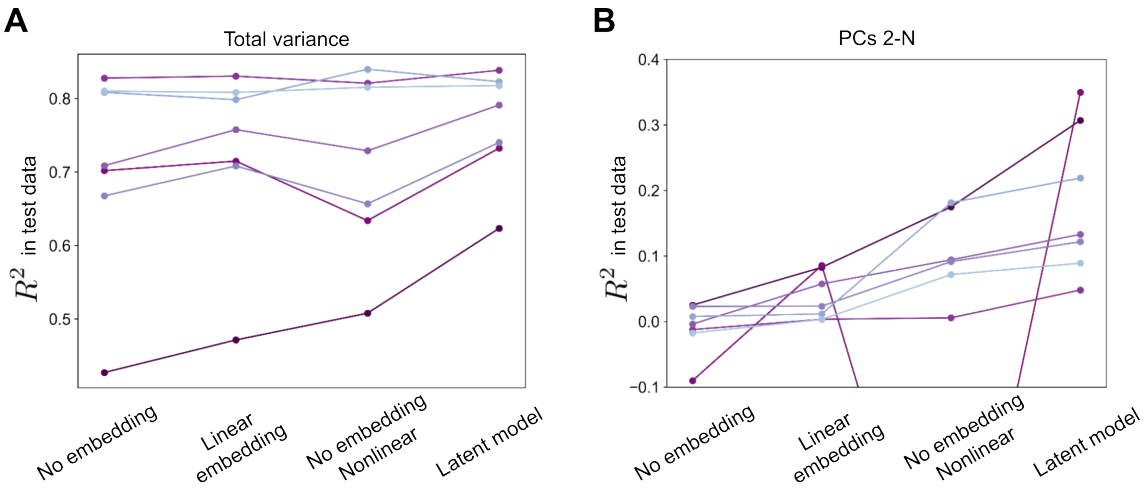

Figure S4: Variance explained in test data for four model categories: "No embedding" (linear pupil regressor), "Linear embedding" (linear mapping from delay embedded pupil), "No embedding, Nonlinear" (nonlinear mapping from pupil regressor), and "Latent model" (nonlinear mapping from delay embedded pupil). Left, total variance explained; right, variance explained along PCs 2-N. Compare also with simulation in Fig. S3C&E.

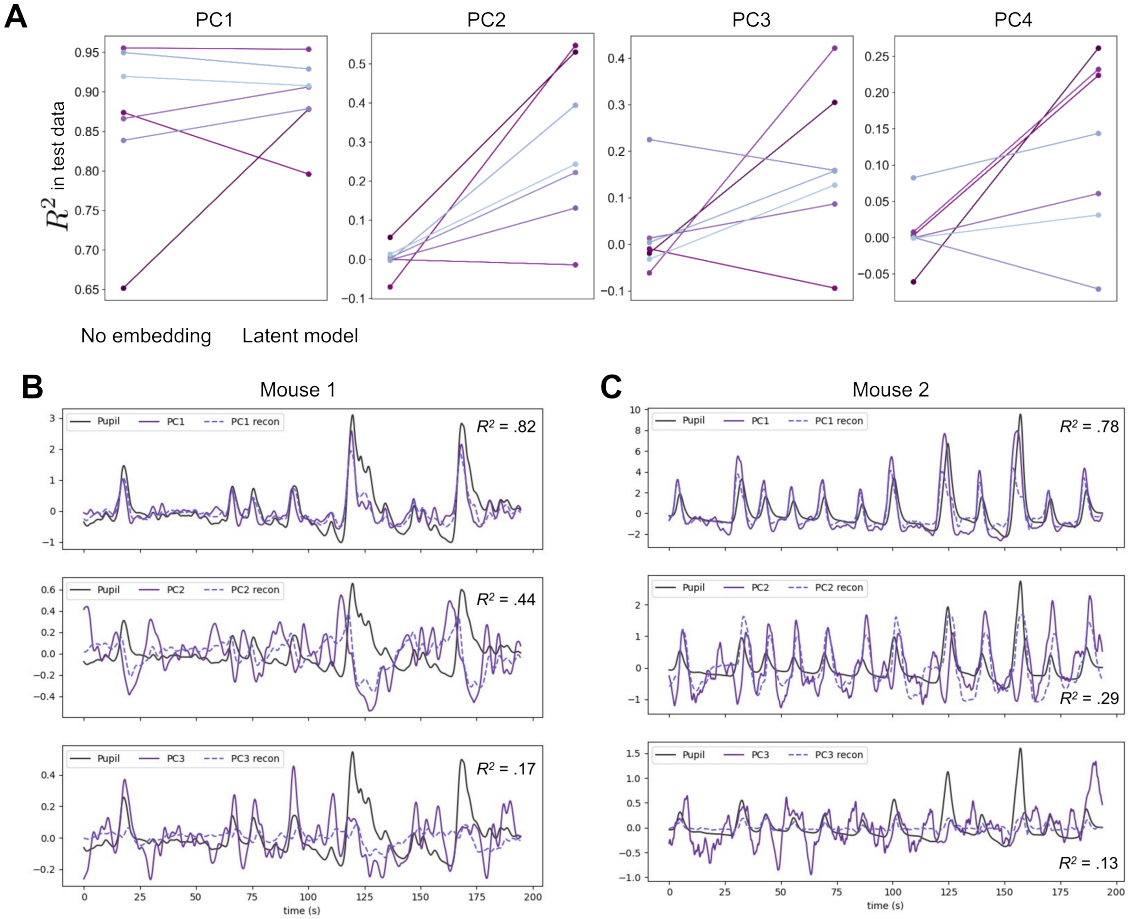

Figure S5: Pupil delay embeddings explain multidimensional variance in widefield calcium images. **A** Variance explained along the dimensions spanned by each of the first four principal components (computed from the training data of each mouse). **B** Principal component time series and their prediction from pupil delay embeddings in test data, shown for two example mice. 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.

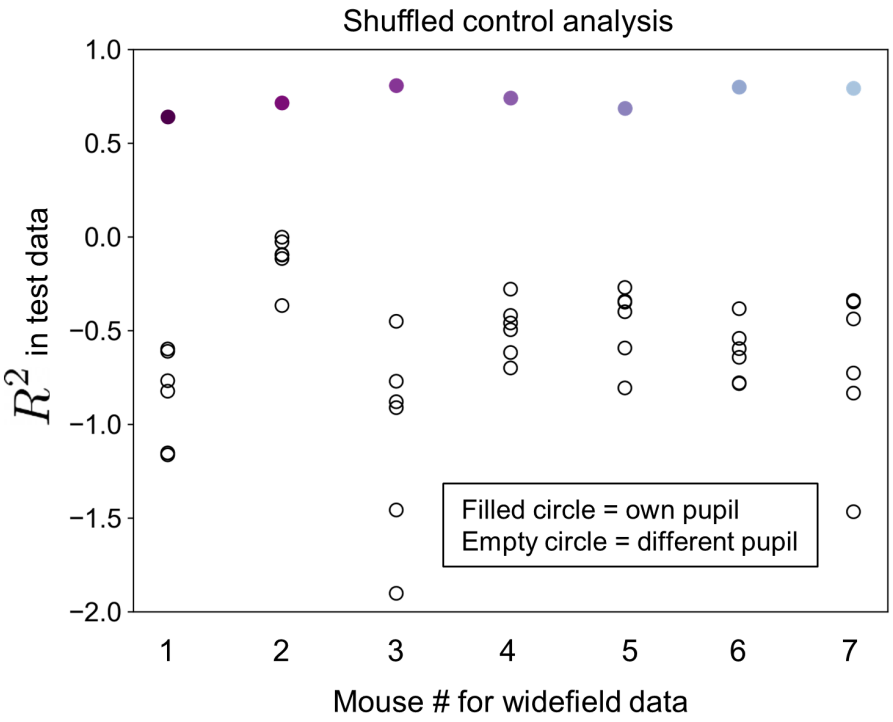

Figure S6: Shuffled control analysis for prediction of widefield calcium data from pupil dynamics (compare with Fig. 2A). For each mouse, we trained seven models to predict the widefield data from pupil dynamics, with each model using the pupil time series of a different mouse. Explained variance (R^2) is reported for each model in held out data. 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). This result clarifies that good performance in held-out data is not trivially obtained with any complex model. (Note that the values of filled circles slightly differ from Fig. 2A as the present analysis did not include an lag adjustment based on the pupil:widefield cross-correlation function.)

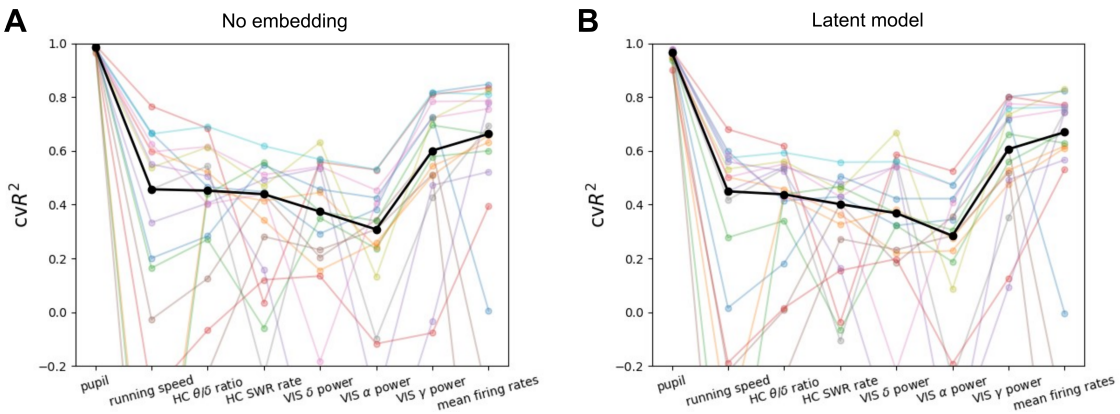

Figure S7: Leave-one-out cross validation analysis in Allen Brain Observatory data. Colored lines correspond to individual mice; black lines reflect median values across mice.