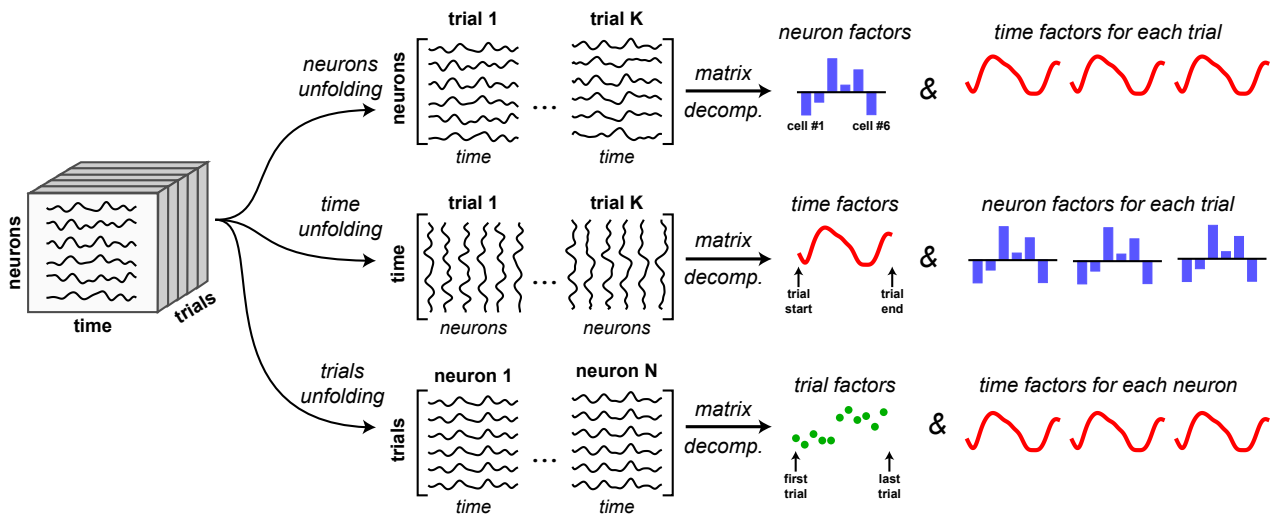# Supplemental Information

Unsupervised Discovery of Demixed,

Low-Dimensional Neural Dynamics across Multiple

Timescales through Tensor Component Analysis

Alex H. Williams, Tony Hyun Kim, Forea Wang, Saurabh Vyas, Stephen I. Ryu, Krishna V. Shenoy, Mark Schnitzer, Tamara G. Kolda, and Surya Ganguli

Figure S1. Related to Figure 2. Illustration of *tensor unfolding* for applying matrix decompositions to tensor datasets. A $N \times T \times K$ dimensional tensor can be reshaped into three different matrices: a "neurons unfolding" with dimensions $N \times TK$, a "time unfolding" with dimensions $T \times NK$, and a "trials unfolding" with dimensions $K \times NT$. Applying PCA or other matrix decomposition methods to each unfolding yields a different set of low-dimensional factors.
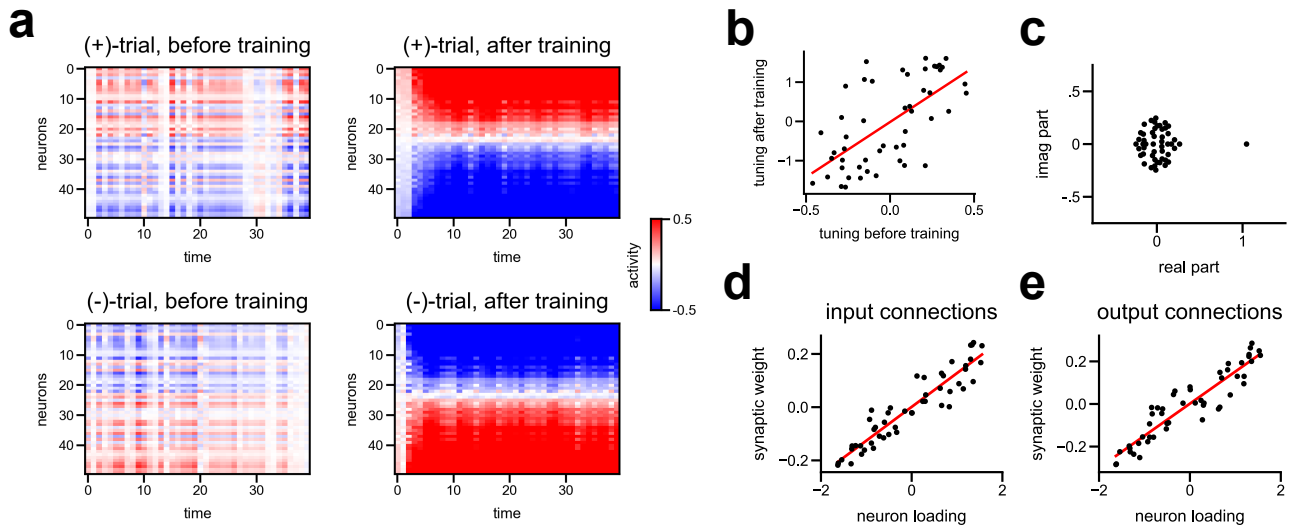
Figure S2. Related to Figure 3. Cell tuning and synaptic connectivity properties in a nonlinear RNN trained on a stimulus discrimination task. **(a)** Activity of all cells on (+)-trials and (-)-trials before and after training. Cells were sorted by the low-dimensional neuron factor, w_n^1, as in Figure 3e. **(b)** Cell tuning quantified as peak activity on (+)-trials minus peak activity on (-)-trials before and after training (averaged over ten trials). Cells with positive tuning scores are (+)-cells, while cells with negative tuning scores are (-)-cells. The initial tuning was positively correlated with final tuning for each cell. **(c)** Eigenvalues of the synaptic connectivity matrix after training. Similar to the solution in linear networks, the connectivity matrix has a single eigenvalue near $1 + 0i$; and all other eigenvalues are small in magnitude. **(d-e)** The neuron factor identified by a 1-component TCA model is positively correlated with the input-to-network synaptic weights **(d)**, and the network-to-output weights **(e)**.
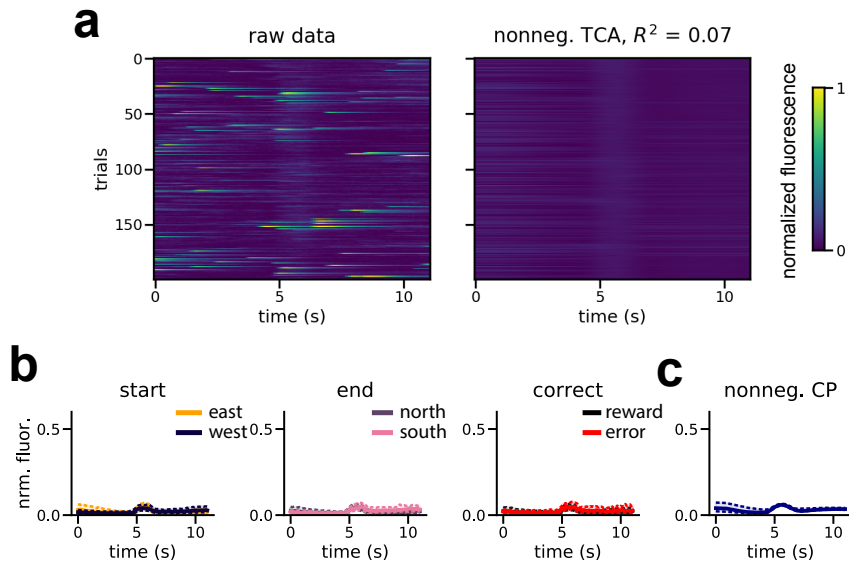
Figure S3. Related to Figure 4. An example cell with low R^2. **(a)** Raster heatmaps showing the cell's fluorescence over the 200 most active trials (left), and the estimate of a 15-component nonnegative TCA model on these trials (right). On a small subset of trials the cell is active, but at variable phases of the trial. Note that on the remaining trials, the cell was hardly active at all (not shown). **(b)** Median fluorescence traces averaged over various task variables (start location, end location, and reward delivery). The cell does not, on average, show a preference for any task variable. Dashed lines denote the first and third quartiles of the fluorescence trace. **(c)** Median estimated fluorescence of the 15-component nonnegative TCA model for this cell. The estimate is closely matched to the median firing rates shown in panel b. Dashed lines denote first and third quartiles.
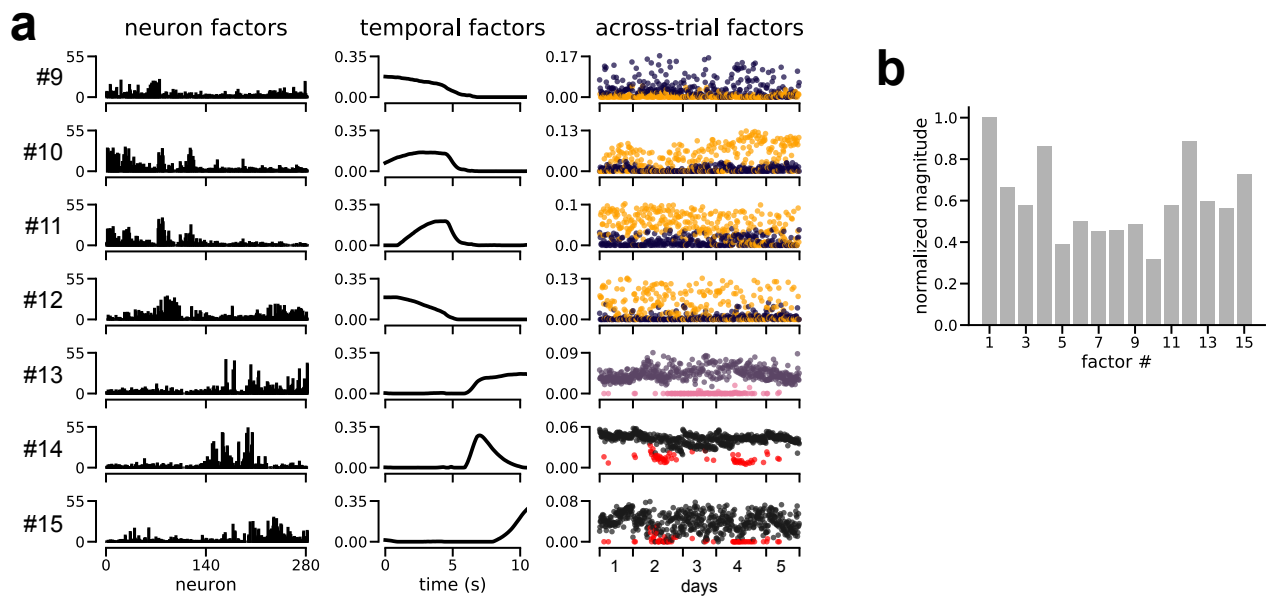
Figure S4. Related to Figure 6. Additional detail on the decomposition of mouse prefrontal cortex dynamics. **(a)** Remaining seven TCA factors from the 15-component decomposition shown in Figure 6. **(b)** The magnitude (Euclidean length) of each factor in the decomposition, a metric analogous to the variance explained by each component (see *Methods*).
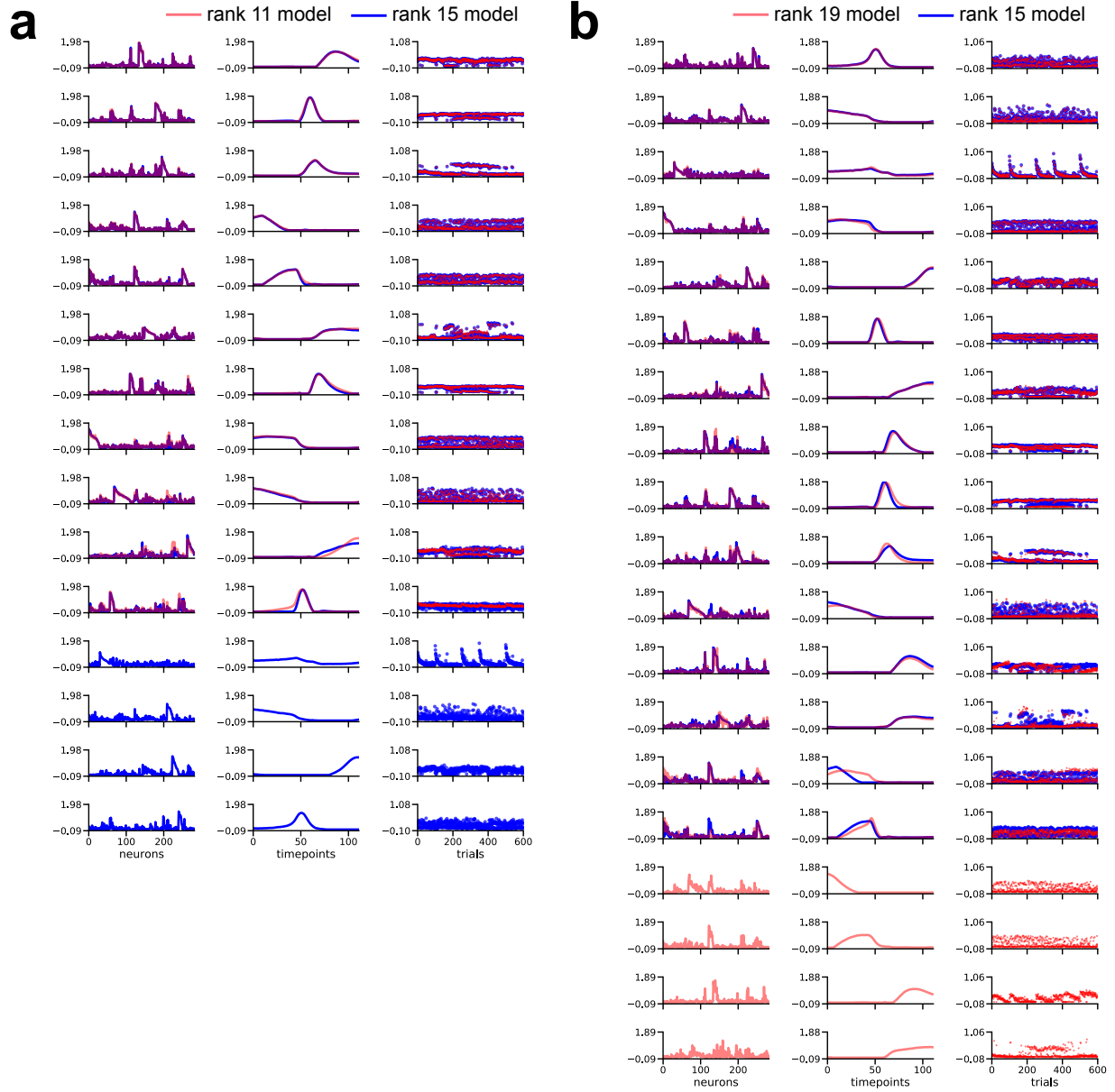
Figure S5. Related to Figure 6. Nonnegative TCA identifies similar factors for different choices of R in the mouse prefrontal cortex dataset. **(a)** The best-fit 15 component model (blue), aligned with the best-fit 11 component model (red). Columns show neural factors (left), temporal factors (middle), and trial factors (right). Rows show components, which are ordered from most similar to least similar. Extra/unmatched components are shown at the bottom. **(b)** Same as panel a except the best-fit 15 component model (blue) is aligned with the best-fit 19 component model (red).
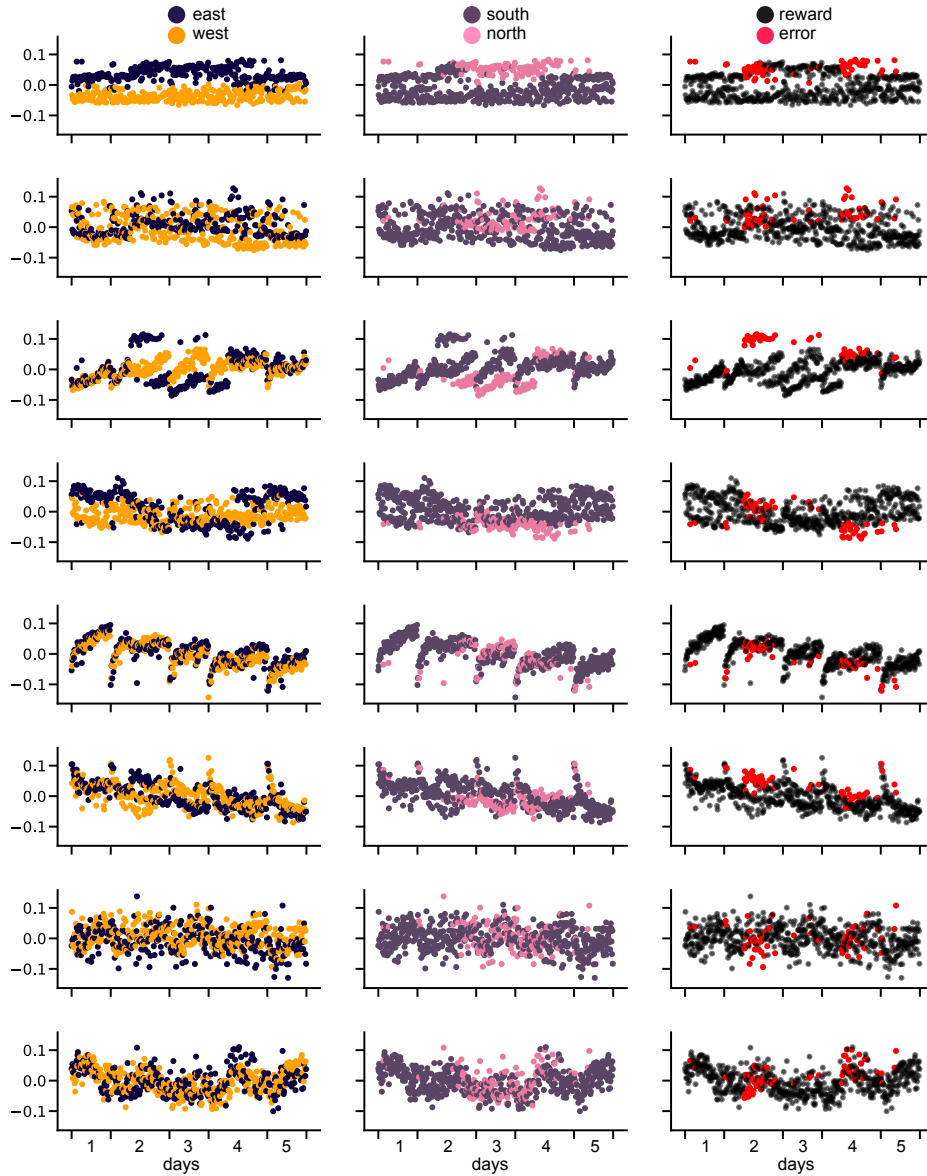
Figure S6. Related to Figure 6. PCA components obtained from the *trials unfolding matrix* (see Figure S1) do not cleanly encode individual task variables. Each row shows a principal component, ordered by variance explained. Each column shows a different coloring of that principal component by a different task variable. With few exceptions (notably the top component), any single coloring does not yield a simple interpretation of the component.
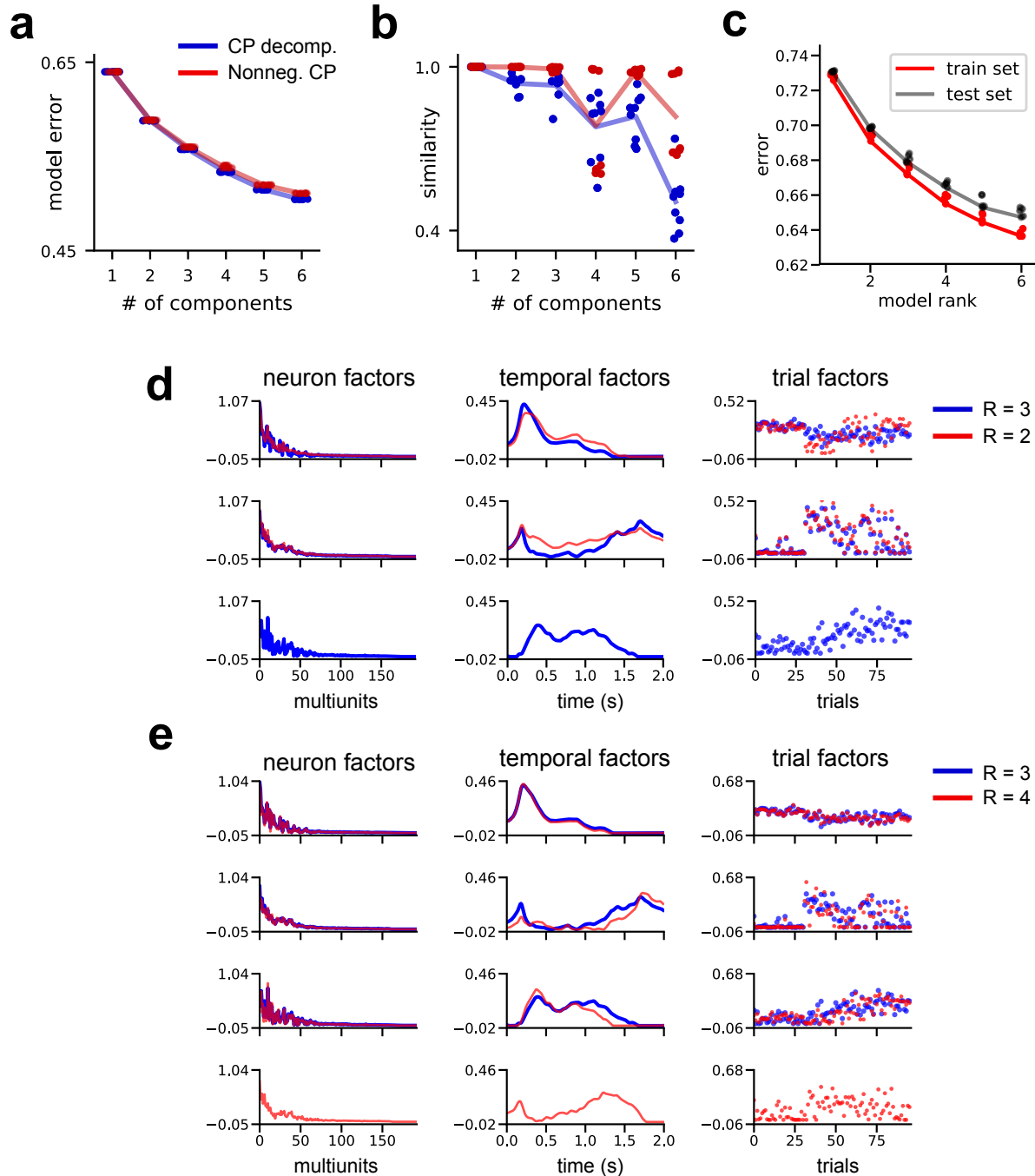
Figure S7. Related to Figure 7. Diagnostic plots for TCA models fit to 45 degree reaches in the primate BMI dataset. **(a)** Error plot for standard (blue) and nonnegative (red) TCA. As elsewhere in this manuscript, each dot denotes a model fit from different initial parameters, demonstrating that neither model got caught in appreciably suboptimal local minima during optimization. Nonnegative decomposition provided similar explanatory power to standard decompositions. **(b)** Similarity plot for standard (blue) and nonnegative (red) CP decompositions. As in other figures, each dot denotes the similarity score between a model and the best-fit model with the same number of components. Nonnegative decomposition had larger similarity scores, suggesting that the latent factors were more reliably identified and less sensitive to initialization. **(c)** Cross-validation of the 45 degree reach dataset, using the procedure described in Figure 5; 80% of the data tensor was heldout at random. **(d)** Comparison of an $R = 2$ nonnegative TCA model (red) with the $R = 3$ model (blue) examined in the main text. The first component of the $R = 2$ model (top row, red) contains features of both the first and third components of the $R = 3$ model (top and bottom rows, blue). **(e)** Comparison of an $R = 4$ nonnegative TCA model (red) with the $R = 3$ model (blue) examined in the main text. The last component of the $R = 4$ model (bottom row, red) ressembles the second component of the $R = 3$ model. Thus, adding an additional component did not add qualitatively new insights.