nature portfolio

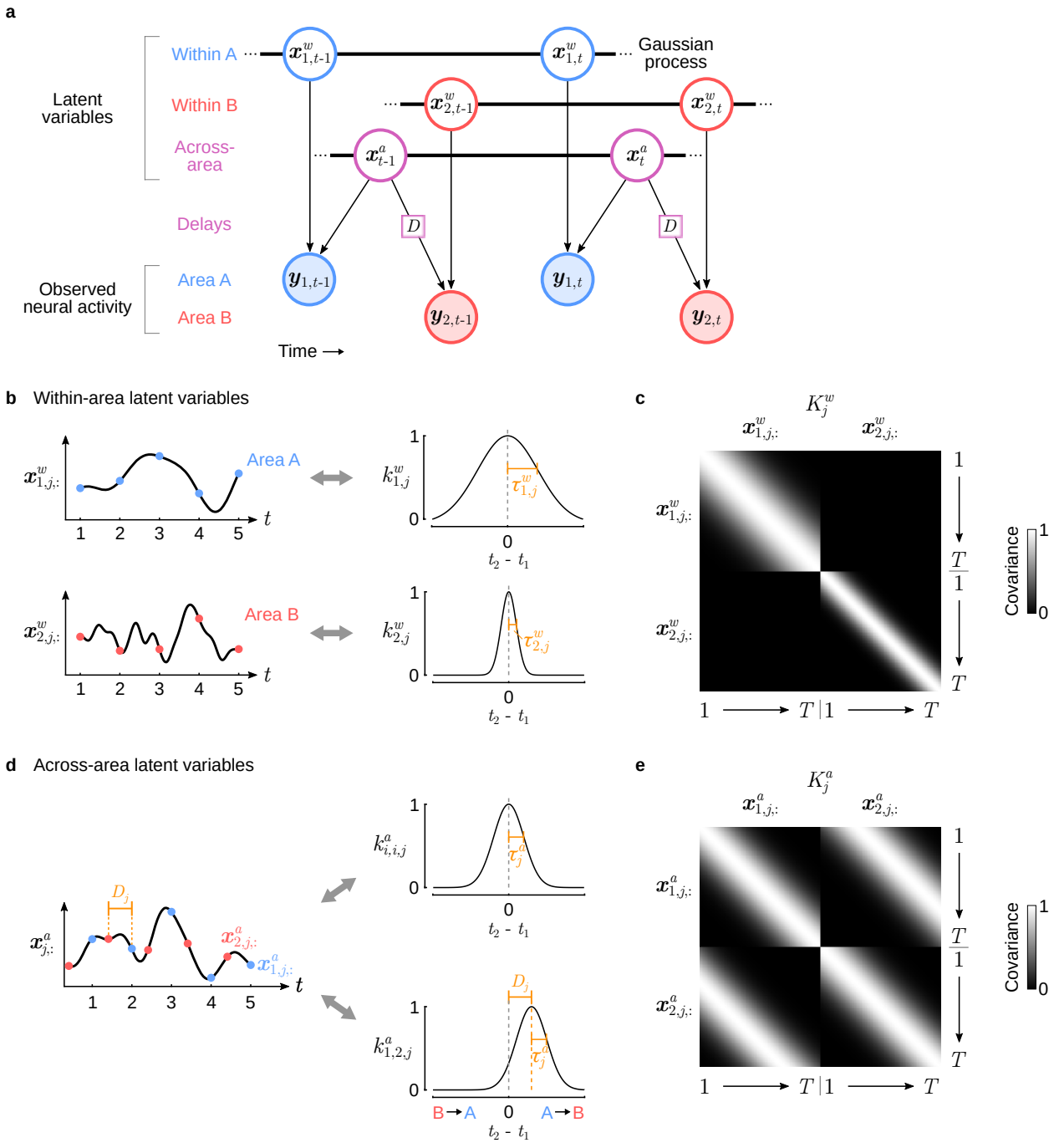
https://doi.org/10.1038/s43588-022-00282-5

Supplementary information

Disentangling the flow of signals between populations of neurons

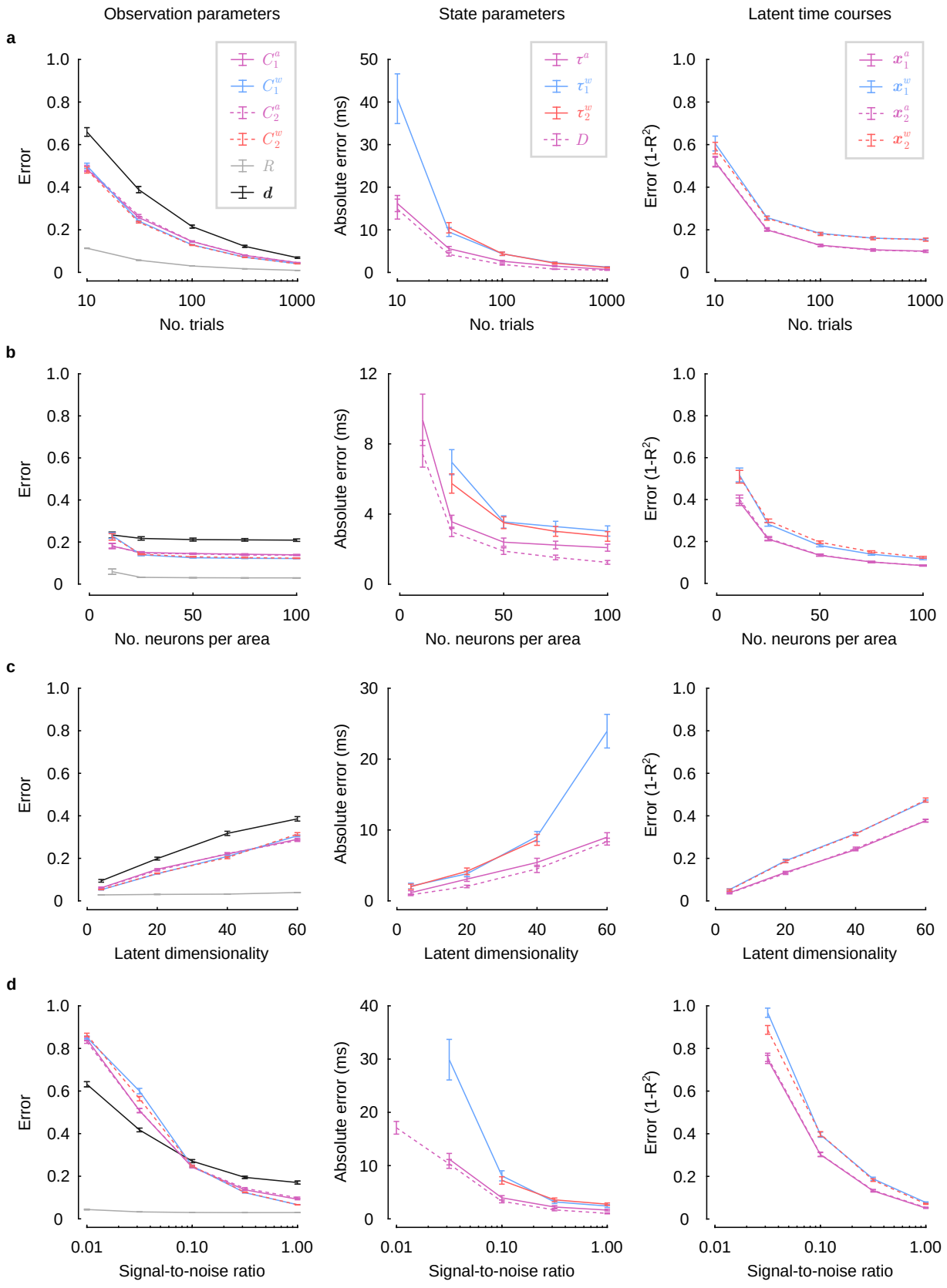
In the format provided by the authors and unedited

Supplementary Information



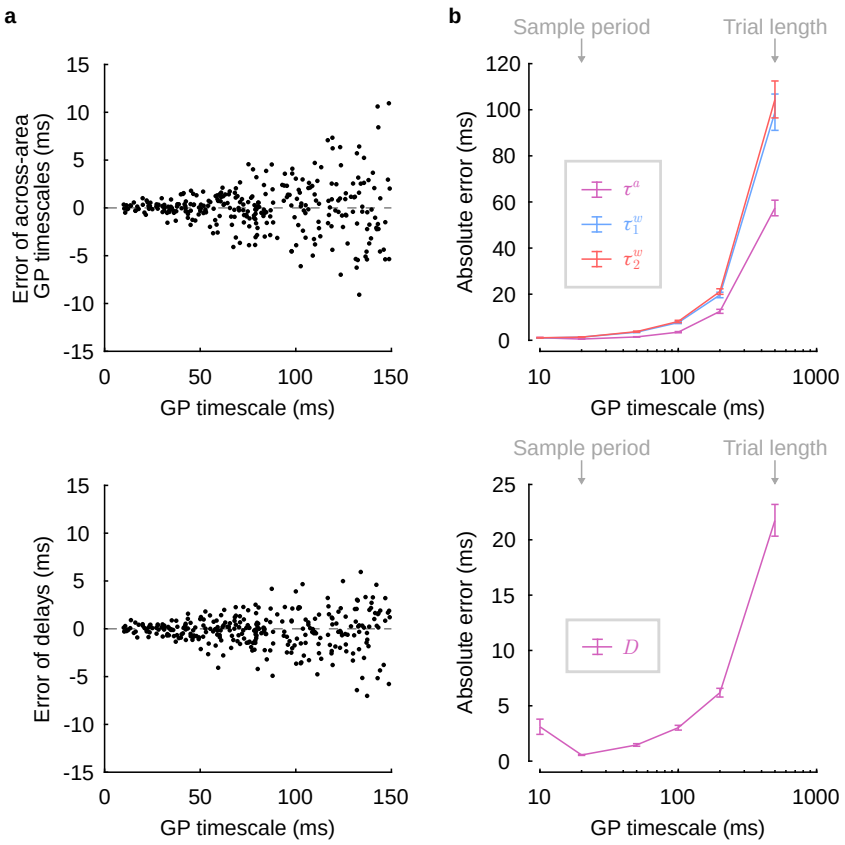
Supplementary Figure 1. DLAG directed graphical model representation, and the use of Gaussian processes in the DLAG state model. (**a**) DLAG directed graphical model representation. Filled circles represent observed variables (i.e., observed neural activity in each area), where $\mathbf{y}_{1,t}$ and $\mathbf{y}_{2,t}$ are the observed neural activity in area A and B, respectively, at time t. Unfilled circles represent latent variables, where $\mathbf{x}_t^a$ are across-area variables at time t; $\mathbf{x}_{1,t}^w$ and $\mathbf{x}_{2,t}^w$ are within-area variables in area A and B, respectively, at time t. D represents the set of relative

time delay parameters between the two areas. Color indicates a variable's or parameter's association with area A (blue), area B (red), or both (magenta). Arrows indicate conditional dependence relationships between variables. In particular, the arrows point from latent variables to observed neural activity, framing DLAG as a generative model. Thick black lines indicate that variables are related in time via a Gaussian process. Here two time steps are shown ($t-1$ and t), and time evolves from left to right. (**b**) Within-area state model. Left column: Within-area time courses (area A: $\mathbf{x}^w_{1,j,:}$, blue points; area B: $\mathbf{x}^w_{2,j,:}$, red points) can be described as a finite number of samples drawn from a Gaussian process (GP) for each area and each j. Right column: The temporal structure of each within-area GP is governed by a covariance function (area A: $k^w_{1,j}$; area B: $k^w_{2,j}$). The squared exponential (SE) function, chosen for the present work, is defined by a timescale parameter ($\tau^w_{1,j}$, $\tau^w_{2,j}$), which controls the width of the covariance kernel, or equivalently, how quickly the latent variable changes over time. (**c**) An example set of within-area GP covariance matrices (K^w_j). The banded structure emerges from the choice of squared exponential function and stationarity of the GP covariance. Note the independence of within-area latent variables across areas: each latent variable has its own characteristic timescale, and cross-covariance terms are all zero. (**d**) Across-area state model. Left column: Like within-area time courses, across-area time courses can also be described as a finite number of samples drawn from a GP. In contrast to the within-area time courses, which are independent across areas, across-area time courses are coupled across areas, drawn from a common GP ($\mathbf{x}^a_{j,:}$). The sampling grid of area A (blue) is shifted by a time delay (D_j) relative to that of area B (red). Right column: The temporal structure of the common GP is governed by a SE covariance function. The width of the auto- and cross-covariances ($k^a_{i,i,j}$ and $k^a_{1,2,j}$, respectively) is controlled by a timescale parameter (τ^a_j). The center of the cross-covariance is controlled by the delay parameter D_j (positive delays: A leads B; negative delays: B leads A). (**e**) An example across-area GP covariance matrix (K^a_j). The banded structure emerges from the choice of squared exponential function and stationarity of the GP covariance. Note the non-zero cross-covariance terms in the off-diagonal blocks of K^a_j: the banded structure is shifted from the diagonal of each off-diagonal block by the delay parameter D_j.


Observation parameters
State parameters
Latent time courses
a
b
c
d
Error
Absolute error (ms)
Error (1-R²)
No. trials
No. neurons per area
Latent dimensionality
Signal-to-noise ratio
C_1^a
C_1^w
C_2^a
C_2^w
R
d
τ^a
τ_1^w
τ_2^w
D
x_1^a
x_1^w
x_2^a
x_2^w

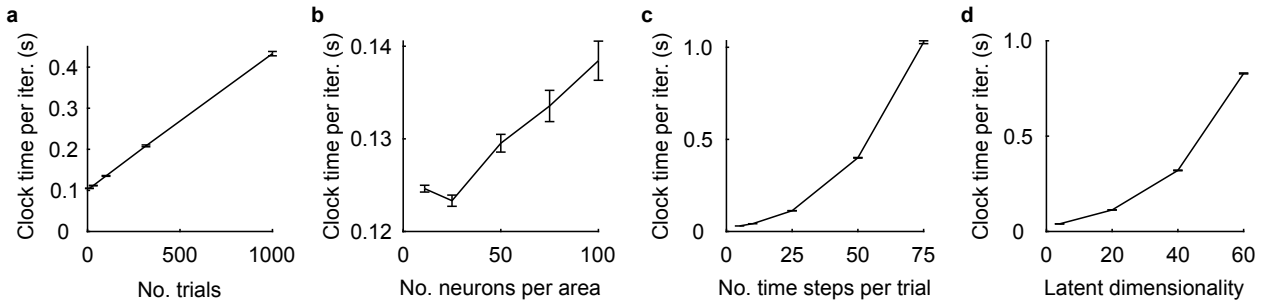
Supplementary Figure 2. DLAG performance as a function of number of trials, number of neurons, latent dimensionality, and signal-to-noise ratio. We sought to characterize DLAG's performance as a function of several data characteristics. For each analysis (panels (a)–(d)), we synthesized 25 datasets (via the DLAG generative model). Unless specified otherwise, the datasets used for each analysis had the following fixed characteristics: $N = 100$ trials; $q_1 = q_2 = 50$ neurons per area; 500 ms trial lengths with 20 ms sampling period (for $T = 25$ samples per trial); latent dimensionalities $p^a = p_1^w = p_2^w = 5$; signal-to-noise ratios $\mathrm{tr}(C_1 C_1^\top)/\mathrm{tr}(R_1) = \mathrm{tr}(C_2 C_2^\top)/\mathrm{tr}(R_2) = 0.3$; GP timescales $\tau^a, \tau_1^w, \tau_2^w \in [10, 150]$ ms; and delays $D \in [-30, 30]$ ms. For each analysis, we varied one of these characteristics to study how it affected DLAG's performance. All panels follow the same plotting conventions: the left column shows the error of observation model parameter estimates (C_1^a: solid magenta; C_2^a: dashed magenta; C_1^w: solid blue; C_2^w: dashed red; R: light gray; $\mathbf{d}$: dark gray); the center column shows the absolute error (in ms) of state model parameter estimates (τ^a: magenta; τ_1^w: blue; τ_2^w: red; D: dashed magenta); the right column shows the error ($1 - R^2$) of latent variable time course estimates ($\mathbf{x}_1^a$: solid magenta; $\mathbf{x}_2^a$: dashed magenta; $\mathbf{x}_1^w$: solid blue; $\mathbf{x}_2^w$: dashed red). (**a**) DLAG performance improves with increasing number of trials. We generated datasets that comprised $N = 1000$ trials. We then took subsets of trials from these datasets, and fit DLAG to increasingly large subsets (sizes equally spaced on a log scale from 10 to 1000 trials). Left: Error bars represent SEM across 25 independent simulated datasets. Center: Error of within-area timescale estimates (τ_2^w) have been omitted for values of 10 trials, where absolute error was 212.1±174.4 ms (mean and SEM across all within-area timescales). Given insufficient statistical power, some GP timescale estimates (likely for latent dimensions that explain little shared variance within an area) become large (i.e., larger than the length of a trial)—to the point where smoothed population activity in the corresponding dimension is effectively constant within a trial. Error bars represent SEM across 125 latent variables. Right: Error bars represent SEM across 25 independent simulated datasets. (**b**) DLAG performance improves with increasing number of neurons (and fixed latent dimensionality). We generated datasets with $q_1 = q_2 = 100$ neurons per area. We then took subsets of neurons from these datasets, and fit DLAG to increasingly large subsets (11, 25, 50, 75, and 100 neurons in each area). Left: Error bars represent SEM across 25 independent simulated datasets. Center: Error of within-area timescale estimates have been omitted for values of 11 neurons per area, where absolute error was 60.1±39.2 ms for τ_1^w and 93.7±46.3 ms for τ_2^w (mean and SEM across all within-area timescales). Error bars represent SEM across 125 latent variables. Right: Error bars represent SEM across 25 independent simulated datasets. (**c**) DLAG performance declines with increasing latent dimensionality (and fixed number of neurons). We considered four settings of across- and within-area dimensionalities ($p^a = p_1^w = p_2^w = 1,\ 5,\ 10,\ 15$). For each setting, we synthesized 25 independent datasets. Here we define the total latent dimensionality (the horizontal axis in each panel) as $2p^a + p_1^w + p_2^w$. Left: Error bars represent SEM across 25 independent simulated datasets. Center: Error of within-area timescale estimates (τ_2^w) have been omitted for values of 60 total latent dimensions, where absolute error was 171.3±91.7 ms (mean and SEM across all within-area timescales). Error bars represent SEM across all across- or within-area latent variables, across all datasets of a given latent dimensionality setting (i.e., across 25, 125, 250, and 375 latent variables for each respective setting). Right: Error bars represent SEM across 25 independent simulated datasets. (**d**) DLAG performance improves with increasing signal-to-noise ratio. We considered five settings for the signal-to-noise ratio (signal-to-noise ratios were the same for both areas; values were spaced equally on a log scale from 0.01 to 1.0). For each setting, we synthesized 25 independent datasets. Left: Error bars represent SEM across 25 independent simulated datasets. Center: Error of GP timescale estimates have been omitted for values of 10^{-2} and $10^{-1.5}$, where absolute errors were greater than 100 ms. Error bars represent

SEM across 125 latent variables. Right: Error of latent time course estimates have been omitted for values of 10^{-2}, where average R^2 values were less than 0 (and hence error values were greater than 1). Error bars represent SEM across 25 independent simulated datasets.

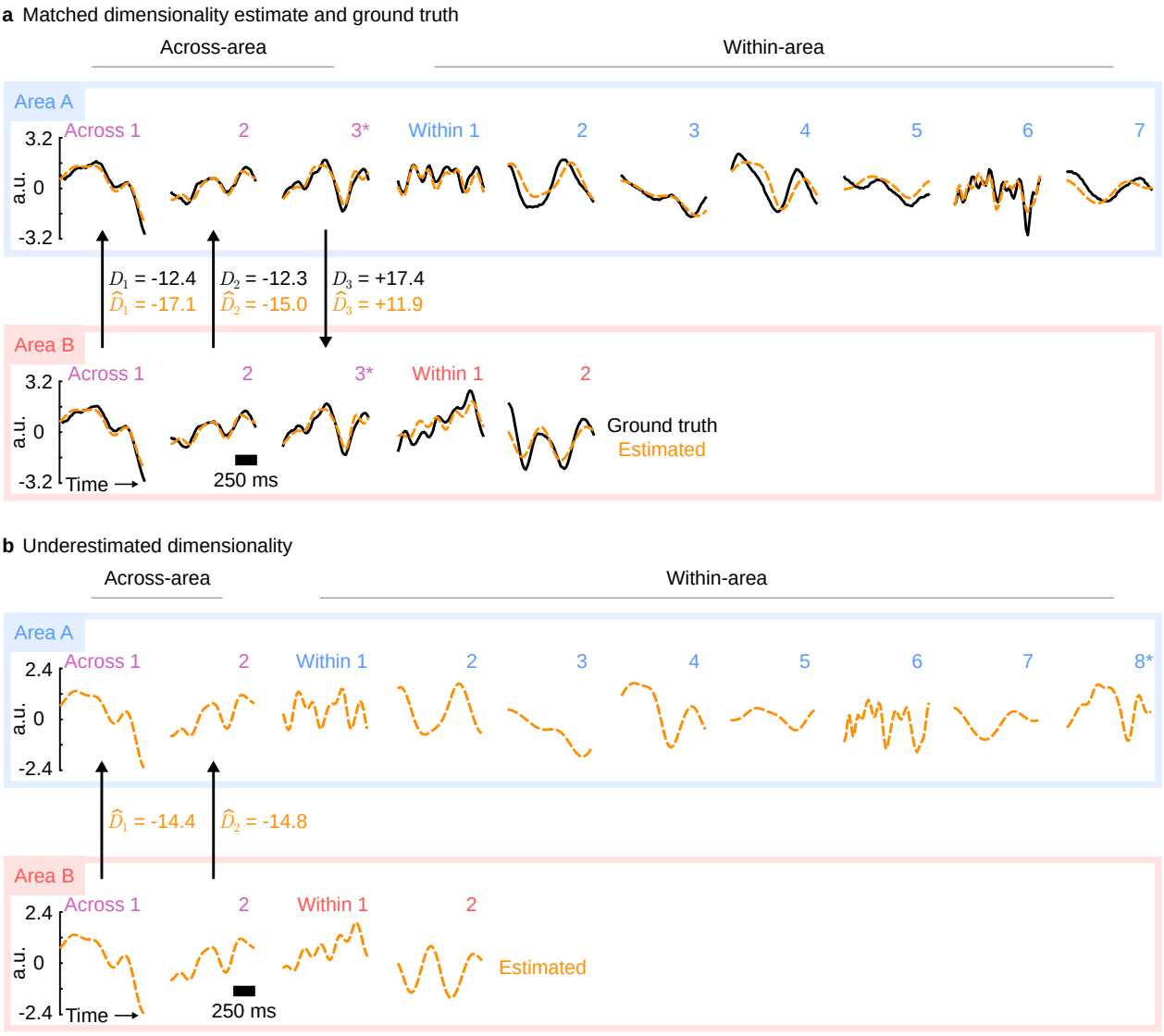


Supplementary Figure 3. Uncertainty of DLAG timescale and delay estimates increases with increasing latent timescale. (**a**) Error (in ms; estimate minus ground truth value) of across-area GP timescale (top) and delay (bottom) estimates for each latent variable shown in Fig. 3c,d. The variance of both GP timescale and delay estimates increases as the underlying ground truth GP timescale increases. For intuition, consider the extreme case of a latent variable whose time course is constant, or equivalently, whose autocovariance function (Supplementary Fig. 1) is flat (i.e., has a very long timescale). Then, a range of DLAG models with any delay and any sufficiently long GP timescale could explain the data equally well, particularly in the presence of noise. (**b**) To verify the trend in (a), we systematically characterized the accuracy of GP timescale and delay parameter estimates as a function of ground truth GP timescale. We synthesized additional datasets (via the DLAG generative model) with the following characteristics: $N = 100$ trials; $q_1 = q_2 = 50$ neurons per area; 500 ms trial lengths with 20 ms sampling period (for $T = 25$ samples per trial); latent dimensionalities $p^a = p_1^w = p_2^w = 5$; signal-to-noise ratios $\mathrm{tr}(C_1 C_1^\top)/\mathrm{tr}(R_1) = \mathrm{tr}(C_2 C_2^\top)/\mathrm{tr}(R_2) = 0.3$; and delays $D \in [-30, 30]$ ms. Each dataset's within- and across-area latent variables were given the same GP timescale; and across 150 datasets, we considered six different timescales (25 datasets synthesized for each timescale), ranging in length from half the sampling period to the length of the trial (10 ms, 20 ms, 50 ms, 100 ms, 200 ms, 500 ms). Top: Absolute error (in ms) of across- and within-area GP timescale estimates increases as underlying GP timescale increases (τ^a: magenta; τ_1^w: blue; τ_2^w: red). Bottom: Absolute error (in ms) of delay parameter estimates increases as underlying GP timescale increases. The lowest error is achieved when GP timescales are equal to the sampling period of observations. For GP timescales larger than the sampling period, the error increases according to the intuition outlined above. For GP timescales less than the sampling period, error increases because a partic-

ular delay can be difficult to estimate if its magnitude is large relative to the corresponding GP timescale: the cross-covariance function (Supplementary Fig. 1) decays quickly enough that observed activity appears uncorrelated across areas in that latent dimension. Error bars represent SEM across 125 latent variables.

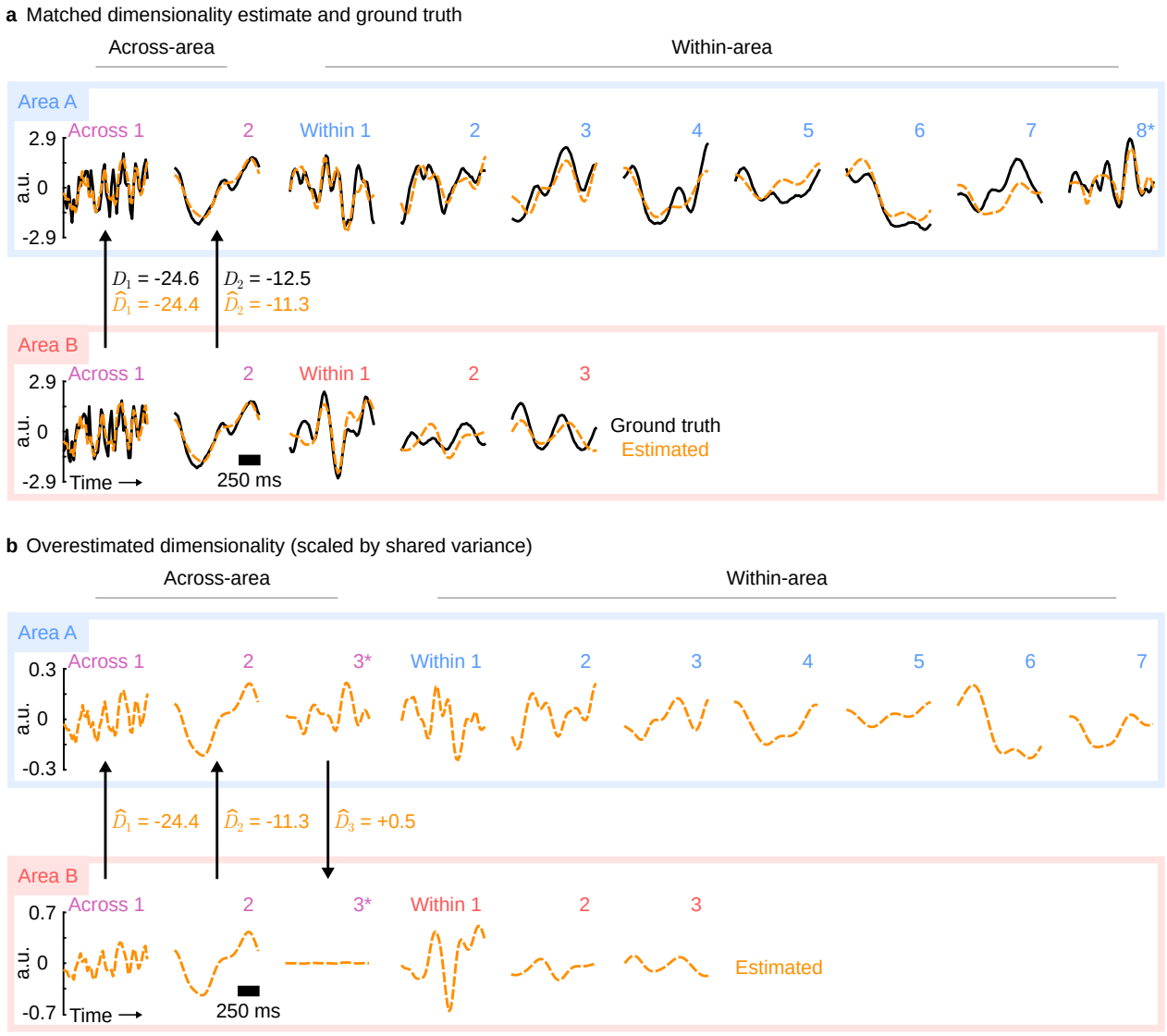


Supplementary Figure 4. DLAG runtime as a function of number of trials, number of neurons, trial length, and latent dimensionality. (**a**) The average clock time (in seconds) per DLAG EM iteration scales (approximately) linearly with the number of trials. These runtime analyses were carried out on synthetic datasets with $q_1 = q_2 = 50$ neurons in each area; $T = 25$ time steps per trial; and latent dimensionalities $p^a = p_1^w = p_2^w = 5$ (total number of latent dimensions $2p^a + p_1^w + p_2^w = 20$). (**b**) The average clock time (in seconds) per DLAG EM iteration scales (approximately) linearly with the number of neurons per area. These runtime analyses were carried out on synthetic datasets with $N = 100$ trials; $T = 25$ time steps per trial; and latent dimensionalities $p^a = p_1^w = p_2^w = 5$ (total number of latent dimensions $2p^a + p_1^w + p_2^w = 20$). (**c**) The average clock time (in seconds) per DLAG EM iteration scales (approximately) quadratically with the number of time steps per trial. Runtime scales quadratically, rather than linearly (as in (a)), because DLAG describes the temporal structure within each trial via Gaussian processes. These runtime analyses were carried out on synthetic datasets with $N = 100$ trials; $q_1 = q_2 = 50$ neurons in each area; and latent dimensionalities $p^a = p_1^w = p_2^w = 5$ (total number of latent dimensions $2p^a + p_1^w + p_2^w = 20$). (**d**) The average clock time (in seconds) per DLAG EM iteration scales (approximately) quadratically with the total number of latent dimensions $(2p^a + p_1^w + p_2^w)$. These runtime analyses were carried out on synthetic datasets with $N = 100$ trials; $q_1 = q_2 = 50$ neurons in each area; and $T = 25$ time steps per trial. In (a)-(d), error bars represent SEM across 25 independent simulated datasets. Results were obtained on a Red Hat Enterprise Linux machine (release 7.9, 64-bit) with 250GB of RAM running Matlab (R2019a), on an Intel Xeon CPU (E5-2695 v3, 2.3 GHz).



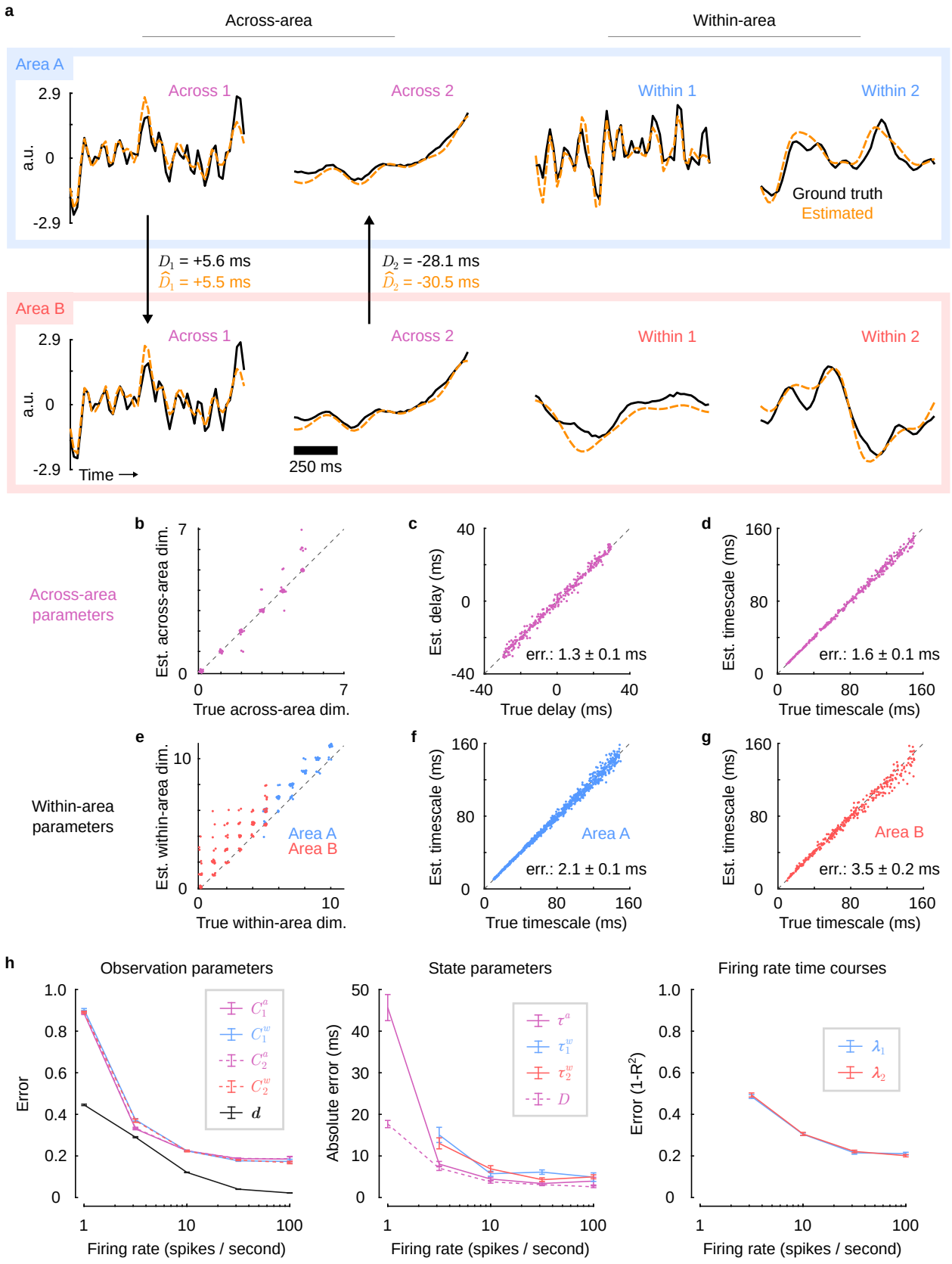
Supplementary Figure 5. DLAG's parameter and latent variable estimates remained stable when dimensionality was underestimated. While the results in Fig. 3b,e suggest that our model selection procedure performs well on realistic-scale synthetic data, we additionally sought to explore the impact of imperfect dimensionality estimates—inevitable in real data—on the estimation and interpretation of DLAG's parameters and latent variables following fitting. With the goal of inducing dimensionality misestimates, we therefore repeated the analyses in Fig. 3b,e with 120 additional datasets generated from the DLAG generative model, but we lowered the signal-to-noise ratio, $\mathrm{tr}(C_i C_i^\top)/\mathrm{tr}(R_i)$, to 0.1 for each area i (compared to 0.3 and 0.2 in area A and area B, respectively, in the original synthetic datasets; see Methods). All other data characteristics remained the same as in the original data. Model selection remained accurate overall: estimated across- and within-area dimensionalities never deviated from the ground truth by more than one (results not shown). Any inaccuracy primarily originated from the initial factor analysis (FA) stage of model selection, rather than the second stage involving DLAG.

Here we present a case study from one of the synthetic datasets described above, in which the total dimensionality of area B was underestimated during the initial factor analysis (FA) model selection stage, and across-area dimensionality was underestimated in the second stage. (**a**) For reference, we first fit a DLAG model with the correct number of within- and across-area latent variables, i.e, no model selection was performed. With real data, we would not have access to this information, but here we use it to understand the scenario in (b). Shown are single-trial latent-variable time course estimates produced by the fitted model along with the ground truth (one example trial shown). Top row / blue box: area A; bottom row / red box: area B. Left: across-area; right: within-area. Orange dashed traces: DLAG estimates; black solid traces: ground truth. a.u.: arbitrary units. Delays reported in ms. Even in the weak-shared variance regime, estimates are qualitatively close to the ground truth. The asterisks ('*') are intended to highlight the third across-area latent variable for each area, which becomes mistaken as a within-area A latent variable when area B's dimensionality is underestimated (see within-area A latent variable 8 in (b)). (**b**) We next consider the model chosen through model selection, as we would with real data. The estimated number of latent variables in area B and the estimated number of across-area variables were each one fewer than the respective ground truth. Shown are single-trial latent-variable time course estimates produced by this model (same trial shown as in (a)). Qualitatively, time course estimates closely match those of the model in (a), in which the correct number of within- and across-area variables was used (compare estimated latent variables with the same index across (a) and (b)). Furthermore, delay estimates are only slightly affected. By inspection, the third across-area latent variable pair (marked by the asterisks in (a)) now appears as the eighth within-area A latent variable (also marked by an asterisk). Note that the ordering of latent variables is arbitrary; we have ordered the latent variables here to facilitate visual illustration.



Supplementary Figure 6. DLAG's parameter and latent variable estimates remained stable when dimensionality was overestimated. Here we present a case study from one of the synthetic datasets described in Supplementary Fig. 5, in which the total dimensionality of area B was overestimated during the initial factor analysis (FA) model selection stage, and across-area dimensionality was overestimated in the second stage. (**a**) For reference, we first fit a DLAG model with the correct number of within- and across-area latent variables, i.e., no model selection was performed. With real data, we would not have access to this information, but here we use it to understand the scenario in (b). Shown are single-trial latent-variable time course estimates produced by the fitted model along with the ground truth. Same conventions as in Supplementary Fig. 5. Even in the weak-shared variance regime, estimates are qualitatively close to the ground truth. The asterisk ('*') is intended to highlight the eighth within-area A latent variable, which becomes mistaken as an across-area variable when area B's dimensionality is overestimated (see across-area variable 3 in (b)). (**b**) We next consider the model chosen through model selection, as we would with real data. The estimated number of latent variables in area B and the estimated number of across-area variables were each one more than the respective ground truth. Shown

are single-trial latent-variable time course estimates produced by this model (same trial shown as in (a)). Qualitatively, time course estimates closely match those of the model in (a), in which the correct number of within- and across-area variables was used (compare estimated latent variables with the same index across (a) and (b)). By inspection, the eighth within-area A latent variable (marked by the asterisk in (a)) now appears as the third across-area latent variable (also marked by asterisks). This phenomenon is straightforward to diagnose: here, we have additionally scaled each latent variable by the fraction of shared variance it explains within its respective area (see Methods; same convention as in Fig. 5). The third across-area latent variable explains little shared variance in area B, consistent with the ground truth. Note that the ordering of latent variables is arbitrary; we have ordered the latent variables here to facilitate visual illustration.


a
Across-area
Within-area
Area A
Across 1
Across 2
Within 1
Within 2
2.9
0
-2.9
a.u.
Ground truth
Estimated
D_1 = +5.6 ms
$\widehat{D}_1$ = +5.5 ms
D_2 = -28.1 ms
$\widehat{D}_2$ = -30.5 ms
Area B
Time
250 ms
b
Across-area parameters
Est. across-area dim.
True across-area dim.
c
Est. delay (ms)
True delay (ms)
err.: 1.3 ± 0.1 ms
d
Est. timescale (ms)
True timescale (ms)
err.: 1.6 ± 0.1 ms
Within-area parameters
e
Est. within-area dim.
True within-area dim.
Area A
Area B
f
err.: 2.1 ± 0.1 ms
g
err.: 3.5 ± 0.2 ms
h
Observation parameters
State parameters
Firing rate time courses
Error
Absolute error (ms)
Error (1-R²)
Firing rate (spikes / second)

Supplementary Figure 7. DLAG accurately estimates within- and across-area time courses and their parameters in synthetic data generated by a linear-nonlinear-Poisson model. We sought to understand how the results in Fig. 3 might change if we applied DLAG to synthetic data in which the linear and Gaussian assumptions of the DLAG observation model, equations (1) and (2), are violated. Toward that end, we generated additional synthetic datasets from the following linear-nonlinear-Poisson (LNP) generative model. For a given dataset, on each trial, we generated within- and across-area latent variable time courses according to the DLAG state model, equations (3)–(8). Hence each latent variable time course followed a Gaussian process (GP) with squared exponential (SE) covariance function, and across-area latent variables included time delays across areas.

For area i with q_i neurons, we then generated neural firing rates, $\boldsymbol{\lambda}_{i,t} \in \mathbb{R}^{q_i}$, during time bin t of width Δ according to the following model:

$$\boldsymbol{\lambda}_{i,t} = \log\left(1 + \exp\left(C_i^a \mathbf{x}_{i,:,t}^a + C_i^w \mathbf{x}_{i,:,t}^w + \mathbf{d}_i\right)\right) \cdot \Delta$$

The function $\log(1 + \exp(\cdot))$ is the commonly used softplus function (applied element-wise to its arguments), a smooth analogue of the rectified linear function. The parameters $C_i^a \in \mathbb{R}^{q_i \times p^a}$, $C_i^w \in \mathbb{R}^{q_i \times p_i^w}$, and $\mathbf{d}_i \in \mathbb{R}^{q_i}$ have similar interpretations as in equations (1) and (2) of the DLAG observation model. We then generated observed spike counts for neuron j in area i during time bin t, $y_{i,j,t}$, according to a Poisson distribution with rate parameter $\lambda_{i,j,t}$ (the j^{th} element of $\boldsymbol{\lambda}_{i,t}$):

$$y_{i,j,t} \mid \mathbf{x}_{i,:,t}^a, \mathbf{x}_{i,:,t}^w \sim \text{Poisson}(\lambda_{i,j,t})$$

Note that this generative model can be interpreted as describing nonlinear interactions across areas since the conditional distributions $P(\mathbf{y}_{2,t} \mid \mathbf{y}_{1,t})$ and $P(\mathbf{y}_{1,t} \mid \mathbf{y}_{2,t})$ describe nonlinear relationships between the observed neural activity in each area, $\mathbf{y}_{1,t}$ and $\mathbf{y}_{2,t}$.

As we did for the synthetic datasets underlying Fig. 3 (see Methods), we generated synthetic datasets from the LNP generative model that were informed by experimental recordings. For all datasets, we chose the numbers of neurons in each area based on our V1-V2 recordings (area A: $q_1 = 80$; area B: $q_2 = 20$). We set the combined total dimensionality in each area to representative values (area A: $p^a + p_1^w = 10$; area B: $p^a + p_2^w = 5$), but varied the relative number of within- and across-area latent variables across datasets. Generating 20 datasets at each of six configurations ($p^a = 0, \ldots, 5$; $p_1^w = 5, \ldots, 10$; $p_2^w = 0, \ldots, 5$) resulted in a total of 120 independent datasets.

We generated the mean parameter for each area i, $\mathbf{d}_i$, so that the distribution of mean firing rates over time and trials was qualitatively similar to typical mean firing rate distributions encountered in V1 and V2 recordings. Specifically, we drew each element of $\mathbf{d}_i$ from an exponential distribution with mean 20 spikes/second and 10 spikes/second in area A and area B, respectively. To ensure that the synthetic datasets exhibited realistic noise levels, we manually tuned the loading matrix parameters for each area, C_i, so that the signal-to-noise ratios according to DLAG model estimates, $\text{tr}(\hat{C}_i \hat{C}_i^\top)/\text{tr}(\hat{R}_i)$, were similar to those encountered in V1 and V2 (0.3 in area A; 0.2 in area B).

Finally, we drew all timescales ($\{\tau_j^a\}_{j=1}^{p^a}$, $\{\tau_{1,j}^w\}_{j=1}^{p_1^w}$, $\{\tau_{2,j}^w\}_{j=1}^{p_2^w}$) uniformly from $U(\tau_{\min}, \tau_{\max})$, with $\tau_{\min} = 10$ ms and $\tau_{\max} = 150$ ms. We drew all delays ($\{D_1, \ldots, D_{p^a}\}$) uniformly from $U(D_{\min}, D_{\max})$, with $D_{\min} = -30$ ms and $D_{\max} = +30$ ms. All Gaussian process noise variances ($\{(\sigma_j^a)^2\}_{j=1}^{p^a}$,

$\{(\sigma^w_{1,j})^2\}_{j=1}^{p^w_1}$, $\{(\sigma^w_{2,j})^2\}_{j=1}^{p^w_2}\}$) were fixed at 10^{-3}. With all model parameters specified, we then generated $N = 100$ independent and identically distributed trials according to the LNP generative model described above. Each trial was $1,000$ ms in length, comprising spike counts in $T = 50$ time bins of width 20 ms, the same spike count bin width used to analyze the V1-V2 recordings. Panels (a), (c), (d), (f), and (g) demonstrate DLAG's ability to estimate the ground truth latent variable time courses and parameters of the LNP generative models when the correct within- and across-area dimensionalities are assumed. Panels (b) and (e) show the results of estimating across- and within-area dimensionalitities from the data.

(**a**) Single-trial latent-variable time course estimates for a representative synthetic dataset. Same conventions as in Fig. 3a. Across all synthetic datasets for which across- or within-area dimensionality was non-zero (across: 100 datasets; within A: 120 datasets; within B: 100 datasets), mean accuracy (R^2) of firing rate estimation was as follows: area A – 0.81; area B – 0.76 (all SEM values less than 0.01). Similarly, mean accuracy of subspace (loading matrix) estimation was as follows: C^a_1 – 0.77; C^a_2 – 0.83; C^w_1 – 0.79; C^w_2 – 0.83 (where a value of 1 implies that the ground truth is fully captured by estimates; all SEM values less than 0.01). (**b**) Across-area dimensionality estimates versus the ground truth for all 120 synthetic datasets. Data points are integer-valued, but randomly jittered to show points that overlap. (**c**) Delay estimates versus the ground truth. Displayed error ('err.') indicates mean absolute error and SEM reported across 300 across-area variables. (**d**) Across-area Gaussian process (GP) timescale estimates versus the ground truth. Displayed error ('err.') indicates mean absolute error and SEM reported across 300 across-area variables. (**e**) Within-area dimensionality estimates versus the ground truth for all 120 synthetic datasets (blue: within-area A; red: within-area B). Data points are integer-valued, but randomly jittered to show points that overlap. (**f**) Within-area A GP timescale estimates versus the ground truth. Displayed error ('err.') indicates mean absolute error and SEM reported across 900 within-area variables in area A. (**g**) Within-area B GP timescale estimates versus the ground truth. Displayed error ('err.') indicates mean absolute error and SEM reported across 300 within-area variables in area B.

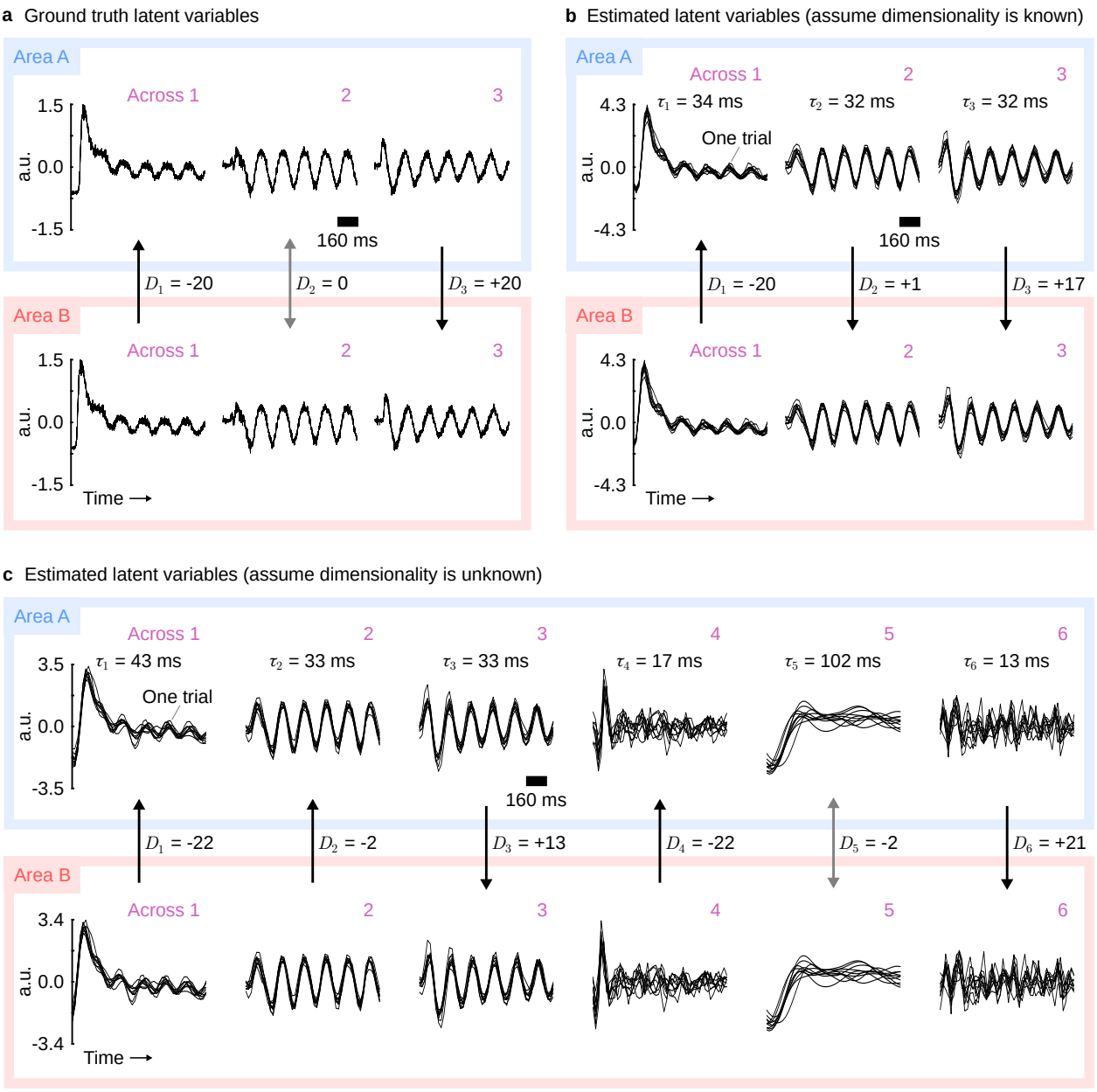
Overall, these results suggest that, for firing rates similar to those encountered in the experimental recordings we consider in this work, DLAG is largely robust when the neural activity is not generated according to the linear and Gaussian assumptions of the DLAG observation model. Across the neuronal populations, firing rates are sufficiently high that neural activity is essentially operating in the linear regime of the softplus function, and a Gaussian noise model can still suffice for Poisson-distributed spike counts (we explore low-firing-rate regimes in panel (h)).

The LNP-generated activity appears to have the greatest impact on the estimation of across- and within-area dimensionalities, shown in panels (b) and (e). During the first stage of our model selection procedure, the optimal factor analysis (FA) dimensionality was larger than the ground truth in at least one area in 115 of 120 datasets. Consequently, estimated within-area dimensionalities also tend to be higher than the ground truth. Interestingly, across-area dimensionality estimates remained highly accurate, matching the ground truth in 107 of 120 datasets (across-area latent activity is shared among a larger number of neurons, leading to greater statistical power). We have already explored the consequences of misestimates of dimensionality in Supplementary Fig. 5 and Supplementary Fig. 6; those results still hold here. Quantifying the shared variance explained by each latent variable (see Methods) provides safeguards against the overestimation of dimensionality.

(**h**) DLAG performance remains stable over a range of realistic firing rates. To probe the limits of DLAG's performance as a function of firing rate, we synthesized additional datasets from the LNP generative model defined above, with the following characteristics: $N = 100$ trials; $q_1 = q_2 = 50$ neurons per area; 500 ms trial lengths with 20 ms spike count bin widths (for $T = 25$ bins per trial); latent dimensionalities $p^a = p_1^w = p_2^w = 5$; GP timescales $\tau^a, \tau_1^w, \tau_2^w \in [10, 150]$ ms; and delays $D \in [-30, 30]$ ms. We systematically varied the mean parameter, $\mathbf{d}$, of the models used to generate each dataset (equally spaced on a log scale from 1 to 100 spikes/second). All neurons had the same mean parameter value, so that mean firing rates over time and trials were nearly the same for all neurons. We manually tuned the loading matrix parameters for each area, C_i, so that the signal-to-noise ratios according to DLAG model estimates, $\mathrm{tr}(\hat{C}_i \hat{C}_i^\top)/\mathrm{tr}(\hat{R}_i)$, were no greater than 0.2 for all firing rate settings. For Poisson-distributed spike counts, the estimated signal-to-noise ratio is inextricably linked to firing rate: in the lowest firing rate setting, 1.0 spikes/second, estimated signal-to-noise ratios were about 0.04. We generated 25 independent datasets for each firing rate setting.

Left: Error of observation model parameter estimates decreases with increasing firing rate, d (C_1^a: solid magenta; C_2^a: dashed magenta; C_1^w: solid blue; C_2^w: dashed red; $\mathbf{d}$: dark gray). Error bars represent SEM across 25 independent simulated datasets. Center: Absolute error (in ms) of state model parameter estimates decreases as firing rate increases (τ^a: magenta; τ_1^w: blue; τ_2^w: red; D: dashed magenta). Error of within-area timescale estimates have been omitted for values of 1 spike/second, where absolute error was 685±236 ms for τ_1^w and 1089±339 ms for τ_2^w (mean and SEM across all within-area timescales). Given insufficient statistical power, some GP timescale estimates (likely for latent dimensions that explain little shared variance within an area) become large (i.e., larger than the length of a trial)—to the point where smoothed population activity in the corresponding dimension is effectively constant within a trial. Error bars represent SEM across 125 latent variables. Right: Error ($1 - R^2$) of firing rate time course estimates decreases as mean firing rate increases ($\boldsymbol{\lambda}_1$: blue; $\boldsymbol{\lambda}_2$: red). Error values have been omitted for values of 1 spike/second, where R^2 values were less than 0 (and hence error values were greater than 1). Error bars represent SEM across 25 independent simulated datasets.

Overall, the smooth degradation of performance as mean firing rates decrease is an expected trend: neural activity increasingly inhabits the nonlinear regime of the softplus function, and DLAG's Gaussian noise model becomes a poorer description of the Poisson-distributed spike counts. Importantly, however, DLAG's performance remains stable over a wide range of firing rates, from 100 spikes/second (50 spikes/trial) to as low as 3 spikes/second (1.5 spikes/trial).



Supplementary Figure 8. DLAG performance when state model, in addition to observation model, assumptions are violated. We next sought to investigate the effects of violations to DLAG's Gaussian process state model assumptions. We therefore explored a case study in which the latent time courses of the linear-nonlinear-Poisson (LNP) generative model, described in Supplementary Fig. 7, were inspired by the V1-V2 neural recordings, rather than generated via Gaussian processes. We generated ground truth across-area latent time courses as follows. (For simplicity, we did not consider within-area latent variables in this case study.) First, we applied canonical correlation analysis (CCA) to spike trains (i.e., neuronal spikes counted in 1 ms time bins) from the same V1-V2 dataset as analyzed in Fig. 5. Hence the data consisted of 400 trials, each 1280 ms in length. CCA produces two sets of canonical basis vectors (dimensions)—one for V1 and one for V2. We took the top three canonical dimensions in V1, and projected observed V1 spike trains on each trial onto these canonical dimensions. Then, we averaged the projected

activity in each canonical dimension over trials, to produce a single set of trial-averaged "template" time courses. For each template time course, we took activity in a 1000 ms time window—these snippets became the across-area latent time courses for our simulated area A. We then took another 1000 ms snippet from each template, time-shifted relative to the snippets used for area A—these snippets became the time-delayed across-area latent time courses for our simulated area B.

Next, we generated an observed spike train on each simulated trial from the LNP observation model defined in Supplementary Fig. 7. The same latent time courses were used on each trial, hence all sources of trial-to-trial variability in these simulations arise from Poisson-distributed noise that is independent across neurons. Before applying DLAG, we counted spikes in 20 ms time bins, as we did for the V1-V2 recordings. Remaining dataset characteristics were as follows: $N = 100$ trials; $q_1 = q_2 = 50$ neurons per area. We drew each element of the mean parameter for area i, $\mathbf{d}_i$, from an exponential distribution with mean 20 spikes/second (same for area A and area B). We manually tuned the loading matrix parameters for each area, C_i, so that the signal-to-noise ratios according to DLAG model estimates, $\mathrm{tr}(\hat{C}_i\hat{C}_i^\top)/\mathrm{tr}(\hat{R}_i)$, was 0.3 for both areas.

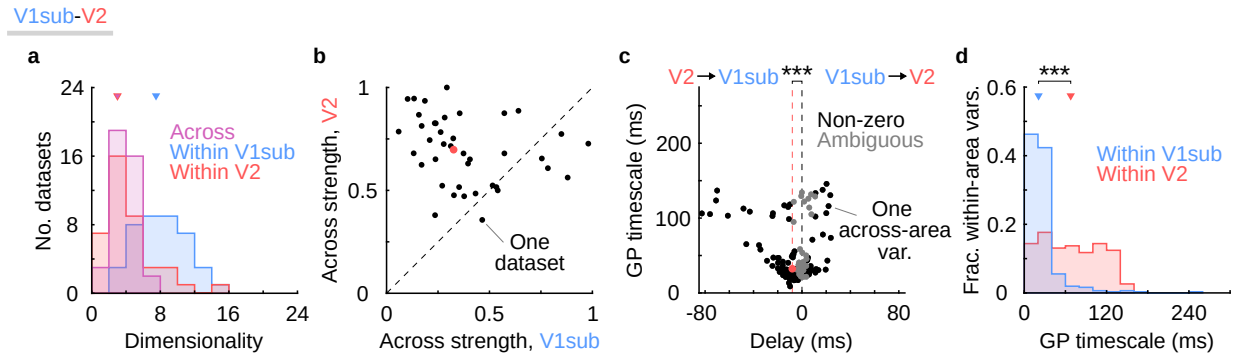
(**a**) Ground truth latent variable time courses. Top row / blue box: area A; bottom row / red box: area B. Across-area variables are paired vertically; vertical arrows point in the direction of signal flow, as defined by the sign of the delay next to each arrow (all delay values are in units of ms). a.u.: arbitrary units. Ground truth latent time courses are the same on every trial. Notice how the assumptions of the DLAG state model (i.e., that latent time courses follow a zero-mean Gaussian process) no longer hold for these simulated data. First, latent time courses are no longer zero-mean. Second, latent time courses no longer covary according to a squared exponential function. For instance, all three across-area latent variable pairs exhibit strong periodic structure. Furthermore, each latent time course comprises multiple timescales: notably, fast transient activity at the beginning of each trial, and slower timescales as the trial progresses.

(**b**) To focus first on the effects these violated assumptions had on DLAG's estimation of latent time courses, without being concerned about model selection, we fit a DLAG model with the same number of across-area latent variables as the ground truth. Each black trace corresponds to one trial; for clarity, only 10 of 100 are shown. To facilitate comparison with panel (c), the estimated GP timescale is displayed for each latent variable (τ_1, τ_2, τ_3). All other conventions are the same as in panel (a). Importantly, DLAG's estimates recapitulate the key qualitative features of the ground truth, including the fast increase in activity at the beginning of each trial (see "Across 1") and the periodic structure throughout each trial. Time delays are also accurately estimated. The latent time courses estimated by DLAG are qualitatively smoother than the ground truth (particularly during the first 60 ms of each trial), a consequence originating from two sources: (1) temporal smoothing via the SE kernel, and (2) counting spikes in 20 ms time bins.

(**c**) Next, we assumed no prior knowledge of the ground truth dimensionality—as would be the case with real neural recordings—and estimated the across-area dimensionality. Interestingly, the optimal across-area dimensionality, selected via cross-validated data log-likelihood (see Methods), was 6, greater than the ground truth value. We investigated the latent time courses extracted by this 6-dimensional model (displayed with the same conventions as in panel (b)).

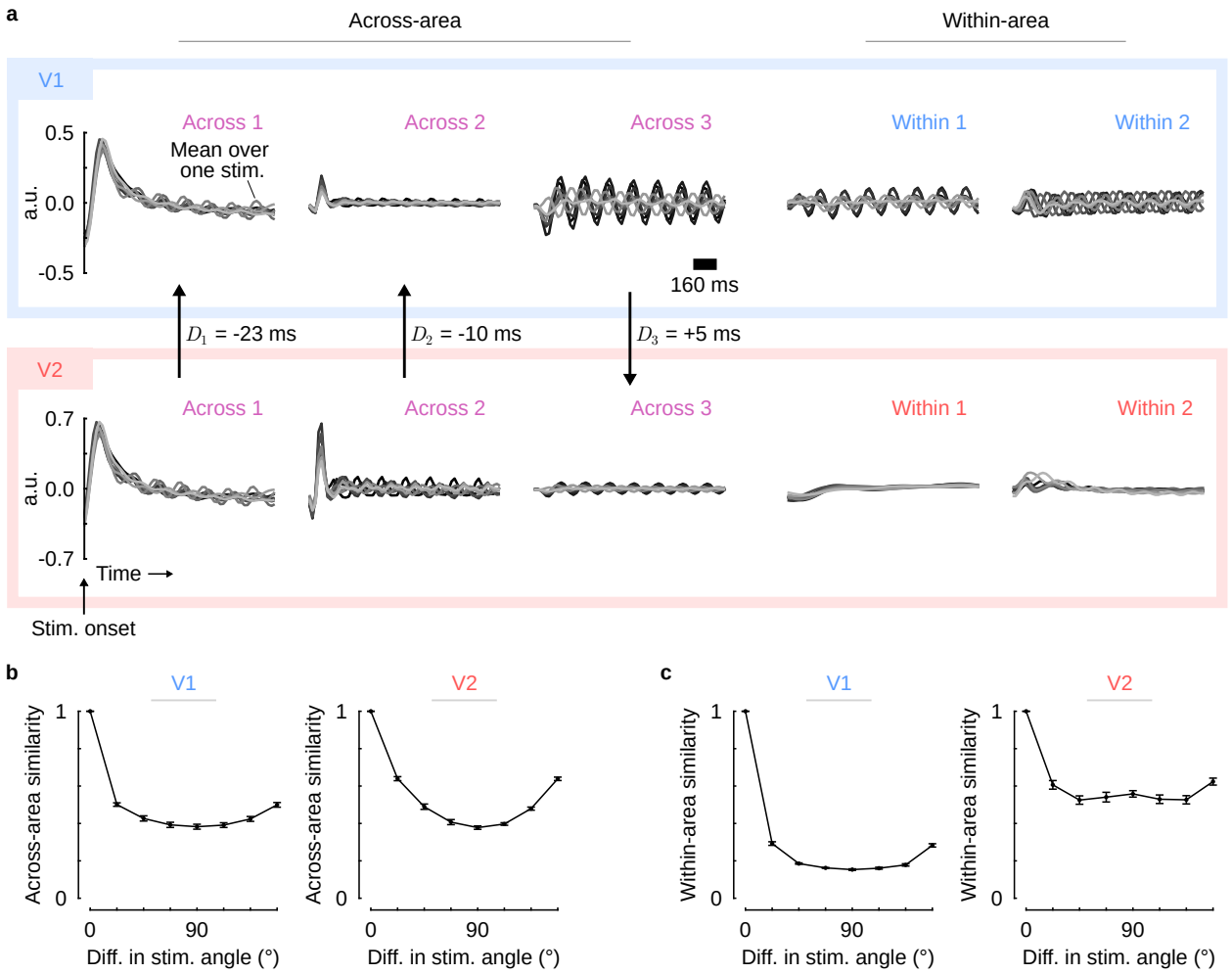
The first three across-area variables still recapitulate the main features of the ground truth. Relative to the 3-dimensional DLAG model (panel (b)), the delay estimates of the 6-dimensional DLAG model differ by a few ms. Close inspection of the first 60 ms of each trial suggests that the first three latent variables of the 6-dimensional DLAG model smooth over the fast transient activity to a greater degree than the 3-dimensional DLAG model. Indeed, Across 1 in (c) has a slightly longer GP timescale (43 ms) than Across 1 in (b) (34 ms).

The remaining latent variables, Across 4–6, are used by DLAG to account for the multiple timescales present in the ground truth. Across 4 combines with Across 1 to account for the fast rise in activity. Across 5 accounts for slower temporal structure throughout the trial, present in all ground truth time courses. Across 6 is periodic with twice the temporal frequency of Across 3, and hence a harmonic signal. Evidence of this phenomenon can be seen in some of the latent variables presented in Fig. 5. We note that we did not rescale latent variable amplitudes here, to best highlight the temporal structure of each latent variable; however, these "extra" latent variables explained little shared variance relative to the first three latent variables (Across 4–6 cumulatively explained only 12% and 9% of the shared variance in area A and in area B, respectively). Still, the model selection results (i.e., that 6 dimensions was deemed optimal) suggest that these extra latent variables do improve DLAG's ability to capture the temporal structure of this simulated neural activity.



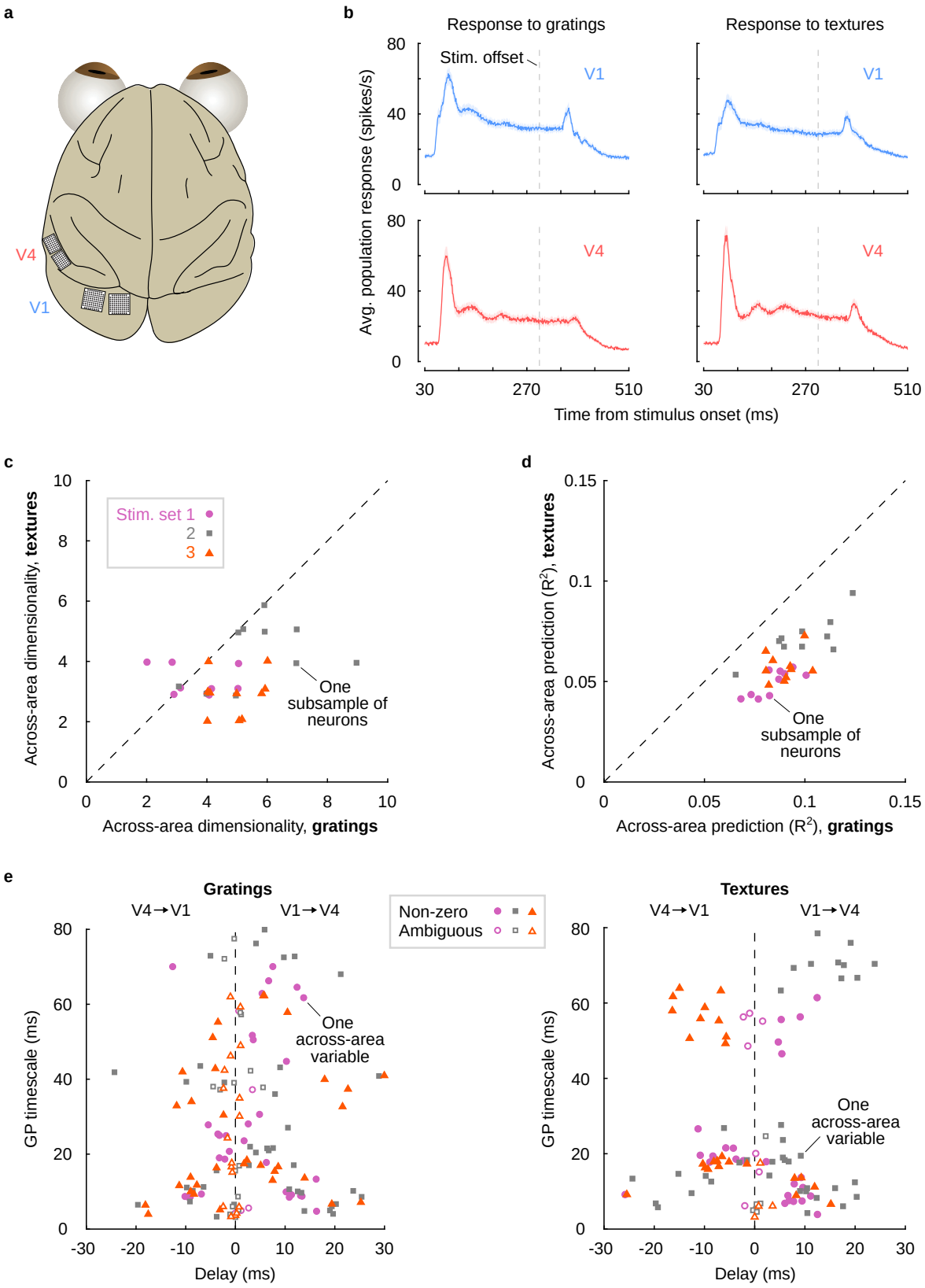
Supplementary Figure 9. V1-V2 results are preserved when V1 is subsampled to match V2 in population size. Same conventions as in Fig. 6. We sought to understand the extent to which the results reported in Fig. 6 were driven by the fact that V1 populations were larger than V2 populations. All else being equal, more neurons allows one to reliably identify more latent dimensions [72]. For each dataset, we thus randomly subsampled the V1 population ('V1sub') to match the size of the V2 population. We then applied DLAG to each subsampled dataset in the same manner as in Fig. 6. (**a**) V1sub-V2 within- and across-area dimensionalities. Compared to Fig. 6, median across-area dimensionality (3) was the same. As a consequence of the smaller population size, median within-V1sub dimensionality (7.5) decreased, but remained higher than median across-area and median within-V2 (3) dimensionalities. Within-V2 dimensionality was 0 in 1 of 40 datasets. (**b**) Fraction of shared variance of each area explained by across-area latent variables in V1sub and in V2. Despite population sizes now being the same, across-area strength is still significantly greater in V2 than in V1sub (median V1sub: 0.33; median V2: 0.70; one-sided paired sign test; $p = 1.8 \times 10^{-4}$), as in Fig. 6. Even after controlling for V1 population size, the within-area dimensionality of V1sub and V2 are not equal. It is possible that the difference in across-area strength seen in (b) is implied by, and therefore redundant with, the difference in within-area dimensionalities seen in (a). Specifically, the weaker across-area strength in V1 relative to V2 might be implied by the greater number of within-V1sub dimensions relative to the number of across-area dimensions. To test this possibility, we recomputed the median across-area strengths for V1 and V2, considering only datasets such that the distributions of within-V1sub and within-V2 dimensionalities were the same. Sixteen datasets remained after this distribution-matching procedure (the 16 datasets for V1 were not necessarily the same 16 datasets as for V2). The medians in V1 and V2 were nearly unchanged (V1sub: 0.33; V2: 0.67). Across-area strengths therefore convey a difference in the properties of V1 versus V2 activity that could not be seen from differences in dimensionality alone. (**c**) Gaussian process (GP) timescale vs. time delay for across-area latent variables. Across all 40 datasets, the delays of 97 of 136 across-area variables were deemed significantly non-zero, and the remaining 39 delays were deemed ambiguous. These values are nearly identical to those reported in Fig. 6. Similarly, delays remained significantly less than zero, representing feedback interactions from V2 to V1sub (median delay across all significantly non-zero across-area variables: -8 ms; '***': one-sided one-sample sign test on 'non-zero' delays, $p = 5.2 \times 10^{-4}$). Among the significantly non-zero delays, 67% were negative. The magnitude of significant negative delays (median: -12 ms) remained greater than the magnitude of significant positive delays (median: +8ms). (**d**) GP timescales for within-area latent variables. GP timescales within V1sub and within V2 are similar to those reported in Fig. 6 (median across 307 within-V1sub latent variables: 21 ms; median across 153 within-V2 latent variables:

68 ms). Furthermore, as in Fig. 6, within-V2 GP timescales are significantly longer than within-V1sub GP timescales ('***': one-sided Wilcoxon rank sum test, $p = 3.5 \times 10^{-29}$).



Supplementary Figure 10. DLAG latent variable time courses and model parameters are sensitive to stimulus condition. (**a**) We first explored how DLAG's estimated latent time courses might change with the orientation angle of grating stimuli. Toward that end, we started with the parameters of the DLAG model fit to the V1-V2 dataset shown in Fig. 5. Then, we used those model parameters to estimate latent time courses on trials involving all eight orientation angles, including the orientation from which the model parameters were fit and the other seven orientations, which were not used for fitting. We then visualized the across- and within-area latent variable time courses, averaged separately over the trials corresponding to each stimulus condition. Left: Across-area time courses. Right: Within-area time courses. Top row / blue box: V1. Bottom row / red box: V2. Each panel corresponds to a latent variable time course. All time courses are aligned to stimulus onset. a.u.: arbitrary units. Each trace represents an average over 400 trials, corresponding to one of eight orientation angles. Gray shading distinguishes stimulus conditions. The ordering of latent variables and all other conventions are the same as in Fig. 5a. The amplitude and phase of the periodic components present in each latent variable reflect the differences in the orientation angle of the drifting gratings. (**b**) We next asked how the across-area subspaces, defined for area i by the loading matrix parameter C_i^a, changed as a function of orientation angle. For each V1-V2 recording session, we computed the subspace similarity ($1 - e_{\text{sub}}$, where e_{sub} is defined in equation (9)) between estimated C_i^a for each pair of grating orientations. To summarize the 8×8 pairwise comparisons in each session, we considered only the relative

differences in orientation angle between datasets (from 0° to 157.5°). We computed the average subspace similarity, across all eight orientations, at each relative orientation angle difference. We then averaged these curves of subspace similarity as a function of relative orientation angle difference across sessions. Left: V1 across-area subspace similarity versus relative difference in orientation angle. Right: Same as the left, for V2. Error bars represent SEM over five sessions, each with eight orientations. Notice the periodic structure of the subspace similarity, consistent with the periodic nature of the oriented grating stimulus: the more similar the orientations, the more similar the identified subspaces. (**c**) Same as (b), but comparing within-area subspaces, defined for area i by the loading matrix parameter C_i^w. Notice again the periodic structure of the subspace similarity, consistent with the periodic nature of the oriented grating stimulus.


a
V4
V1
b
Response to gratings
Response to textures
Stim. offset
V1
V4
Avg. population response (spikes/s)
Time from stimulus onset (ms)
c
Stim. set 1
2
3
Across-area dimensionality, textures
Across-area dimensionality, gratings
One subsample of neurons
d
Across-area prediction (R²), textures
Across-area prediction (R²), gratings
One subsample of neurons
e
Gratings
Textures
V4→V1
V1→V4
Non-zero
Ambiguous
GP timescale (ms)
Delay (ms)
One across-area variable

Supplementary Figure 11. DLAG shows that V1-V4 interactions depend on the type of visual stimulus presented.

We used DLAG to study interactions between a second pair of brain regions (visual areas V1 and V4) in an awake animal. In particular, we sought to explore if DLAG, when used to study V1-V4 interactions, was sensitive to the type of stimulus presented: oriented gratings versus naturalistic textures. Previous work has shown that responsivity to higher order statistics of visual stimuli develops gradually along the ventral visual stream. V2 and V4 respond to the higher order statistics present in textures, whereas V1 does not—selective primarily to the spectral content of textures [73, 74].

To better understand the effect of stimulus complexity on inter-areal communication, we recorded simultaneous V1 and V4 population responses to gratings and textures while an awake animal was passively fixating. Animal procedures and recording details have been described in previous work [41, 75]. Briefly, one male adult cynomolgus macaque was trained to maintain fixation on a small spot ($0.2^\circ \times 0.2^\circ$, 80 cd/m^2) on a gray background (40 cd/m^2) within a 1.4° diameter fixation window. Eye-position was monitored using a video tracking system (Eyelink II, SR research, ON, Canada) with a sampling rate of 500 Hz. Stimuli were presented on a calibrated monitor 64 cm away from the animal (1400×1050 pixel resolution; 100 Hz refresh rate).

(**a**) After training, Utah arrays (0.4 mm spacing; 1 mm electrode length, Blackrock, UT) were implanted in V1 and V4: two 96 channel arrays in V1 and two 48 channel arrays in V4 (see [41]). All procedures were approved by the IACUC of the Albert Einstein College of Medicine. We targeted the arrays to have matching retinotopic locations in V1 and V4 by relying on anatomical markers and previous mapping studies. Receptive fields were in the lower right visual hemifield and largely overlapping for V1 and V4 populations (see [41]). Extracellular voltage signals were amplified and band-pass filtered between 250 and 7.5 kHz using commercial acquisition software (Blackrock Microsystems, UT and Grapevine, Ripple, UT). Voltage snippets that exceeded a user-defined threshold were digitized and sorted offline.

Visual stimuli and task contingencies were presented using custom openGL software (Expo: `http://sites.google.com/a/nyu.edu/expo`). During recording sessions, two sets of stimuli were presented: a set of sinusoidal gratings and a set of naturalistic textures, which included noise stimuli whose spectra were matched to that of a texture. Sets of gratings included four full contrast stimuli, comprising two spatial frequencies one octave apart (1.2-2.4 cyc/$^\circ$) and two orientations 90° apart (e.g., 1.2 cyc/$^\circ$, 45°; 2.4 cyc/$^\circ$, 45°; 1.2 cyc/$^\circ$, 135°; 2.4 cyc/$^\circ$, 135°). Sets of textures included six stimuli (four naturalistic texture stimuli, two spectrally matched noise stimuli), generated as follows. Two textures were selected from the Multiband Texture Database (`http://multibandtexture.recherche.usherbrooke.ca/original_brodatz.html`) and Salzburg Texture Image Database (`https://wavelab.at/sources/STex`). The two textures were first down sampled to 256×256 pixels and matched in contrast. Then two distinct samples (each 512×512 pixels in size) were synthesized for each texture using the Portilla-Simoncelli algorithm [76]. One sample of spectrally matched noise was synthesized for each of the two textures. All stimuli were presented in a 4.7° square aperture.

Trials began with the animal fixating on a small spot in the center of the screen. After a delay of 300 ms, a random sequence of two stimuli, both from either the grating set or the texture set,

appeared on the screen. Each stimulus presentation lasted for 300 ms. The inter stimulus interval was 400 ms (gray screen). After the second stimulus presentation, the animal had to maintain fixation for an additional 300 ms (gray screen) and was then positively reinforced with a liquid reward if fixation was maintained throughout the trial. The animal performed on average 1307±15 trials per session. We recorded neural activity for three sessions.

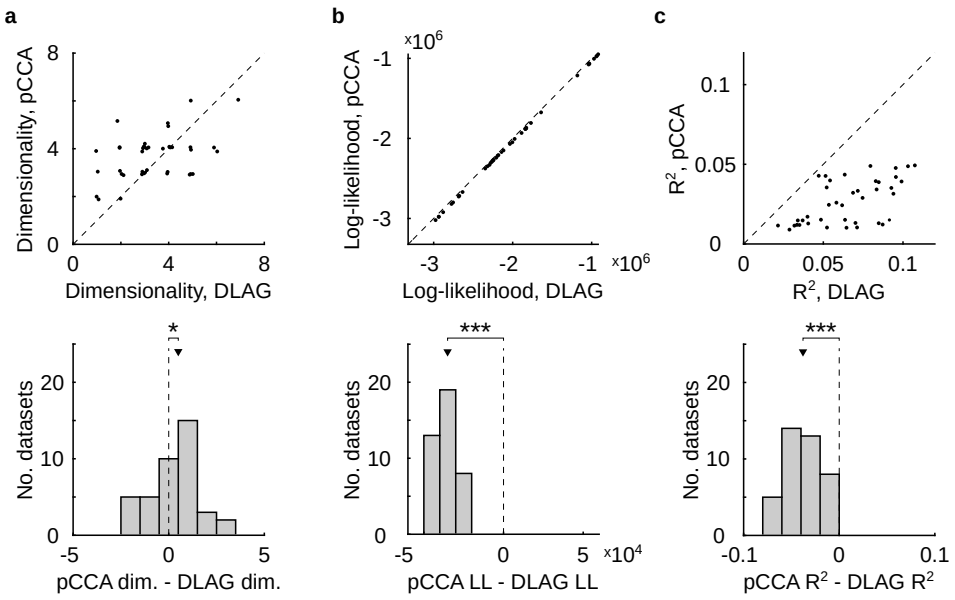
We were interested in observing whether DLAG was sensitive to the presentation of grating versus texture stimuli. Hence for further analysis, we excluded presentations of spectrally matched noise stimuli. As stated above, each trial comprised two stimulus presentation periods: we treated these periods as independent "trials" when applying DLAG. Our analysis included on average 262±4 presentations per stimulus (four grating stimuli, four texture stimuli) per session. For each recording session, we grouped together all trials in which oriented grating stimuli were presented (regardless of orientation or spatial frequency; a "grating stimulus set"), and all trials in which texture stimuli were presented (regardless of texture sample; a "texture stimulus set"). We analyzed 480 ms time windows, from 30 ms after stimulus onset to 210 ms after stimulus offset (hence the analysis time window included some spontaneous neural activity). We counted spikes in 20 ms time bins during this analysis time window.

We analyzed neuronal responses from one V1 array and from one V4 array that showed the greatest visual receptive field overlap with V1. For each recording session, we excluded neurons that fired fewer than 0.5 spikes/second, on average, for any given stimulus condition. We also excluded neurons with a Fano factor greater than 1.6, on average, across all stimulus conditions (Fano factor was computed across trials of one stimulus condition at a time). Following these screening steps, sessions 1, 2, and 3 contained pools of 60, 83, and 77 neurons, respectively, in V1, and pools of 44, 54, and 37 neurons, respectively, in V4. Note that, for each recording session, V1 and V4 neurons were the same across grating and texture stimulus sets. Because we were interested in V1-V4 interactions on timescales within a trial, we subtracted the mean across time bins within each trial from each neuron. This step removed activity that fluctuated on slow timescales from one stimulus presentation to the next.

(**b**) Average population activity in V1 (top row) and V4 (bottom row) in response to an example grating stimulus set (left column) and in response to an example texture stimulus set (right column). These grating and texture stimulus sets correspond to Stimulus Set 2 (gray squares) in panels (c)–(e). Shaded regions indicate ± one SEM, where the mean is taken over peristimulus time histograms (PSTHs) of individual neurons (83 in V1; 54 in V4). The recorded V1 and V4 neurons are the same across the left and right columns.

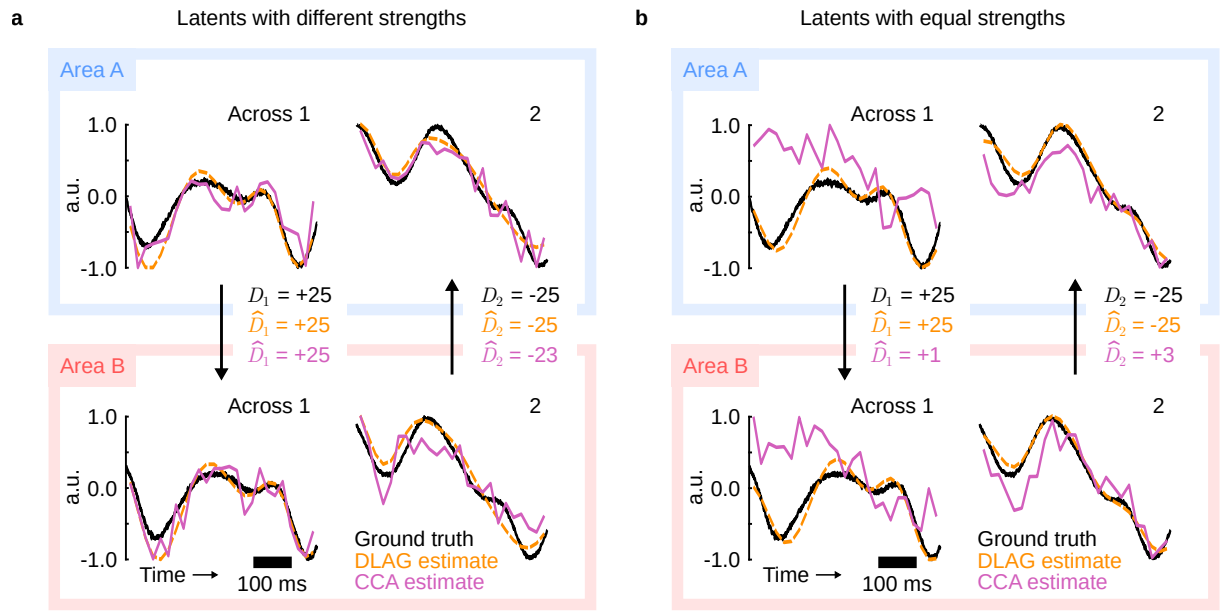
Finally, throughout our analyses, we sought to assess the variability of DLAG's estimates within each recording session and stimulus set. For each recording session, we randomly subsampled 20 V1 neurons and 20 V4 neurons from the overall pool of neurons described above. We repeated this subsampling procedure 10 times (starting from the same overall pool of neurons in V1 and in V4). We then applied DLAG separately to each subsample, resulting in 60 separate analyses across the three recording sessions, each with one grating stimulus set and one texture stimulus set. Importantly, the subsampled V1 and V4 neurons were the same across grating and texture stimulus sets, enabling direct comparison between DLAG models. In (c)–(e), we refer to these paired grating/texture stimulus sets as simply "stimulus sets."

(**c**) V1-V4 across-area dimensionality during the presentation of texture stimuli versus oriented grating stimuli. Each point represents results for a single subsample of V1 and V4 neurons. Data points are integer-valued, but randomly jittered to show points that overlap. In two of three stimulus sets, V1-V4 across-area dimensionality was significantly lower during presentations of texture stimuli than during presentations of oriented grating stimuli (one-sided paired sign test; stimulus set 1, magenta circles: $p = 0.144$; stimulus set 2, gray squares: $p = 0.016$; stimulus set 3, orange triangles: $p = 0.002$). (**d**) Cross-validated across-area prediction (leave-group-out R^2) between V1 and V4 during the presentation of texture stimuli versus oriented grating stimuli. Each point represents results for a single subsample of V1 and V4 neurons. In all three stimulus sets, V1-V4 across-area prediction appears weaker during presentations of texture stimuli than during presentations of oriented grating stimuli (one-sided paired sign test; for all stimulus sets, $p < 0.001$). (**e**) Gaussian process (GP) timescale vs. time delay for across-area latent variables uncovered during presentations of oriented grating stimuli (left) and during presentations of texture stimuli (right). Each point represents one across-area variable. Filled points: across-area latent variables for which the delays were deemed significantly non-zero. Unfilled points: across-area latent variables for which delays were deemed ambiguous (not significantly positive or negative). Relative to grating sets, texture sets exhibited a marked absence of across-area GP timescales in the 30–45 ms range. The time delays for across-area variables with GP timescales in the 45–80 ms range appear to depend on the set of textures presented (points in this range cluster according to texture set).



Supplementary Figure 12. V1-V2 interactions are better described by DLAG than by probabilistic canonical correlation analysis. To demonstrate the advantages of modeling the temporal structure of neuronal interactions within and across areas, we applied probabilistic canonical correlation analysis (pCCA) [42] to the same V1-V2 datasets as in Fig. 6. pCCA is a static dimensionality reduction method that includes across-area latent variables, but not within-area latent variables (see Methods, equations (46) and (47)). (**a**) Comparison of pCCA and DLAG across-area dimensionality estimates. For each of the 40 V1-V2 datasets, we identified the number of pCCA latent variables through K-fold cross-validation (here we chose $K = 4$, as was done for DLAG cross-validation). The pCCA model with the highest cross-validated data likelihood was taken as optimal. Top: Estimated pCCA dimensionality versus estimated DLAG across-area dimensionality. Each data point represents one V1-V2 dataset. Data points are integer-valued, but randomly jittered to show points that overlap. pCCA and DLAG estimates of across-area dimensionality are modestly correlated (Pearson correlation coefficient, $r = 0.48$). Bottom: Distribution of the differences between pCCA and DLAG across-area dimensionality ('dim.') estimates on each dataset. pCCA estimates are slightly higher than DLAG estimates (black triangle indicates the median difference across datasets: 0.5; '*': one-sided paired sign test; $p = 0.0494$). (**b**)–(**c**) DLAG outperforms pCCA according to multiple metrics. On each dataset, we compared the optimal pCCA model to the optimal DLAG model (each selected through cross-validation) via two performance metrics: cross-validated data log-likelihood (LL; b) and cross-validated leave-group-out R^2 (c). See Methods for details. Cross-validated LL offers the most principled comparison, as it is precisely the data log-likelihood that the two probabilistic methods are intended to maximize. However, interpretation of the relative performance differences between methods can be difficult given the scale of LL values. Furthermore, LL values can vary dramatically from dataset to dataset, often by orders of magnitude. Hence leave-group-out R^2 facilitates more intuitive comparison between methods and across datasets, at the expense of a principled characterization of performance within each method's probabilistic framework. Top panels: pCCA performance versus DLAG performance. Each data point represents one V1-V2 dataset. Bottom panels: Distribution of differences between pCCA and DLAG performance on each dataset. DLAG significantly outperforms pCCA across datasets (black triangles indicate the median difference across datasets; '***': one-sided paired sign test; $p < 0.001$). DLAG's better performance can be at-

tributed to multiple differences between the DLAG and pCCA models. First, DLAG includes the addition of low-dimensional within-area latent variables. pCCA models within-area activity via full-rank observation noise covariance matrices (see equation (47)). Fig. 6a suggests that within-area activity in both V1 and in V2 is well-described as low-dimensional. Second, the number of parameters in the DLAG model scales linearly with the number of neurons in each area, whereas the number of parameters in the pCCA model scales quadratically with the number of neurons in each area, lending pCCA to be more prone to overfitting. Third, DLAG accounts for the temporal structure of within- and across-area interactions (using Gaussian processes), whereas pCCA does not. Fourth, DLAG accounts for time delays in across-area interactions, whereas pCCA does not.



Supplementary Figure 13. Canonical correlation analysis (CCA) cannot disentangle signals that are relayed concurrently and with similar strength. Here we leverage simulations to demonstrate where a static method like CCA is unable to disentangle concurrent signaling. In brief, we synthesized two additional datasets from the linear-nonlinear-Poisson (LNP) generative model defined in Supplementary Fig. 7. The two datasets were nearly identical, with one difference: in the first dataset (a), across-area latent variables had different strengths; in the second dataset (b), across-area latent variables had equal strengths. This difference between datasets cannot be seen above, since the amplitudes of latent time courses are normalized.

In detail, we first generated latent time courses for $p^a = 2$ across-area variables. For simplicity, we did not include within-area latent variables. One across-area variable (Across 1) was assigned a delay of +25 ms (so that area A leads area B; observe the relative time-shift in Across 1 between black traces in area A versus area B); the second across-area variable (Across 2) was assigned a delay of −25 ms (so that area B leads area A; observe the relative time-shift in Across 2 between black traces in area A versus area B). Both across-area variables had the same Gaussian process (GP) timescale, 60 ms. In this demonstration, we wanted to isolate the consequences of the CCA model definition from issues like overfitting. We therefore simulated a data-rich scenario by generating $N = 1,000$ independent trials, each 500 ms in length. On each trial, we generated a different set of across-area latent time courses, X_n. Let $X = \{X_1, \ldots, X_N\}$ be the set of latent time courses over all N trials.

For both datasets, we generated spike trains (see Supplementary Fig. 7) at 1 ms resolution for $q_1 = q_2 = 50$ neurons per area from the common set of latent time courses, X. All neurons had the same mean parameter value ($\mathbf{d}$, defined in Supplementary Fig. 7) of 20 spikes/second, so that mean firing rates over time and trials were nearly the same for all neurons. The loading matrix parameters for each area, C_i^a, were manually tuned so that the signal-to-noise ratios according to DLAG model estimates, $\mathrm{tr}(\hat{C}^a{}_i \hat{C}^a{}_i^\top)/\mathrm{tr}(\hat{R}_i)$, were 0.2. We counted spikes in 20 ms time bins, and then fit both a CCA model and a DLAG model to each dataset.

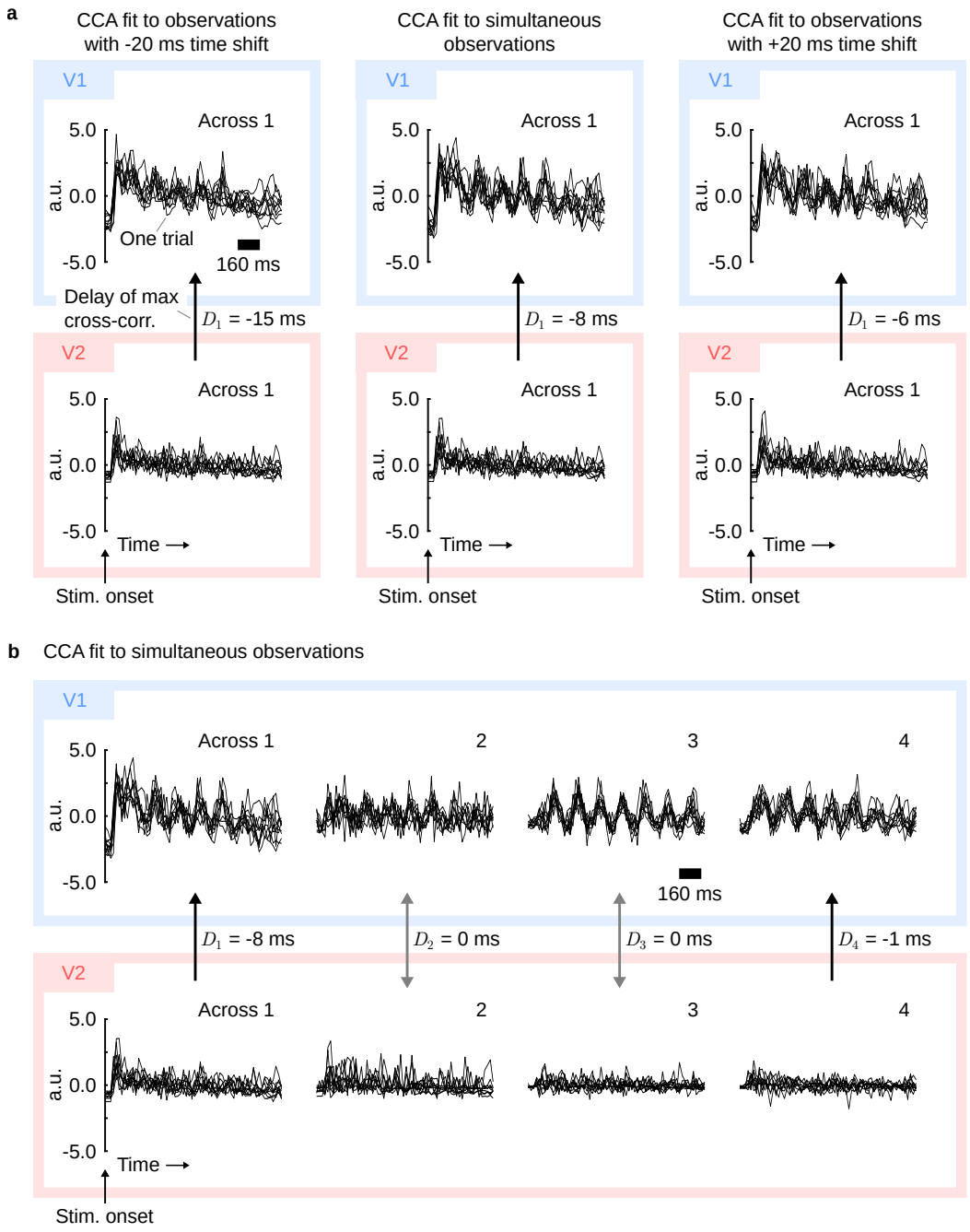
The difference between the two datasets was as follows. For the first dataset, we scaled the columns of C_i^a (for each area i) so that the magnitude of the column associated with the +25 ms latent variable was twice the magnitude of the column associated with the −25 ms latent variable. For the second dataset, we took the same C_i^a that was used for the first dataset, but rescaled the columns of C_i^a (for each area i) so that both columns had equal magnitude. We performed this rescaling such that signal-to-noise ratios remained the same across both datasets. Thus these two datasets allowed us to isolate the effects of the relative strengths of feedforward versus feedback signals on CCA's (and DLAG's) ability to disentangle those signals.

Time delays are not inherently built into the CCA model. To estimate a time delay for each pair of fitted canonical dimensions, we identified the time delay at which projections of area A activity and projections of area B activity had maximum cross-correlation. The cross-correlation function between area A and area B projections was computed with 1 ms resolution, from -40 ms (B leads A) to +40 ms (A leads B). In detail, we first took a fixed window of activity in area A, 420 ms in length, from 40 ms to 460 ms into the trial. For each trial, we counted spikes within this window in 20 ms nonoverlapping time bins, and projected this activity onto each canonical dimension in area A. For area B, we employed a sliding window of length 420 ms, which we advanced in 1 ms increments, from the beginning of the trial to 80 ms into the trial. At each increment, we counted spikes within the window in 20 ms nonoverlapping time bins, and projected this activity (on each trial) onto each canonical dimension in area B. For each canonical pair, we computed the Pearson correlation between the projected area A activity and the projected area B activity. This correlation value gave one element of a cross-correlation function: repeating this procedure at each increment of the sliding window in area B produced a cross-correlation function from -40 ms to +40 ms. We then identified the time delay at which the cross-correlation function for each canonical pair was maximum.

(**a**) Each canonical dimension can reflect a directed interaction if the signals in each direction have different strengths. In general, the first canonical pair returned by CCA is the pair of dimensions along which projections of simultaneously observed activity exhibit the greatest correlation across areas. Projections onto the second canonical pair exhibit the second greatest correlation across areas, and so on. In this dataset, projections of simultaneously observed activity onto Across 1 exhibit greater across-area correlation than do projections onto Across 2, by design. Thus the first and second canonical pairs estimated here indeed reasonably reflect each direction of signal flow. DLAG estimates closely match the ground truth. Top row / blue box: area A; bottom row / red box: area B. Black solid traces: ground truth across-area latent time courses on a representative trial. Orange dashed traces: DLAG estimates. Magenta solid traces: CCA estimates. a.u.: arbitrary units. Black arrows indicate the direction of signal flow between area A and area B, given by the ground truth delay value. Ground truth and estimated delay values (in ms) are shown to the right of each arrow (top, black: ground truth; center, orange: DLAG estimate; bottom, magenta: CCA estimate). Canonical pairs are sorted from left to right, in descending order, based on the value of their canonical correlation.

(**b**) Canonical dimensions reflect a mixture of signals relayed in each direction if those signals have similar strengths. Same conventions as in panel (a). Because the latent variables in this dataset have similar strengths, the canonical pairs do not provide a faithful description of each direction of signal flow. The CCA-estimated time courses and time delays deviate significantly from the ground truth. DLAG estimates still closely match the ground truth.

Overall, these two scenarios demonstrate that CCA can identify directions of signal flow if signals in one direction are dominant (a), but not if signals in both directions have similar strengths (b). DLAG successfully disentangles concurrent signaling in both scenarios.



Supplementary Figure 14. Canonical correlation analysis (CCA) provides a description of V1-V2 signal flow that is qualitatively different from that of DLAG. Here we consider the V1-V2 recordings, and explore the qualitative differences between a static method like CCA and DLAG, particularly in their descriptions of inter-areal signal flow. We thus considered the same V1-V2 dataset as presented in Fig. 5, and studied the projections of V1-V2 neural activity onto the across-area dimensions obtained via CCA.

(**a**) The top canonical variable is dominated by feedback (V2 leads V1) activity, even if CCA is fit to V1-V2 activity with a nominal feedforward (V1 leads V2) time-shift. One approach to us-

ing CCA to identify the direction of inter-areal signal flow was recently proposed in [41]. There, a sliding window scheme was used, in which observations of V2 activity were first time-shifted relative to observations of V1 activity, and then CCA was fit to this time-shifted V1-V2 activity. CCA was fit anew for each incremental advance of the sliding window throughout the course of the trial, thereby producing a different set of canonical dimensions for each relative time shift between V1 and V2 activity. The top canonical dimensions were then studied at various time delays and at various time points throughout the trial to identify periods of feedforward- and feedback-dominated activity.

In Supplementary Fig. 13, we showed that the top canonical dimension—when fit to simultaneous observations—reflects either the dominant direction of interaction (Supplementary Fig. 13a) or a mixture of signals relayed in both directions (Supplementary Fig. 13b). Could one tease apart concurrent feedforward and feedback signals by instead fitting the top canonical dimension to time-shifted V1-V2 activity, as in [41]? One might expect, for example, that a feedforward interaction becomes dominant in V1-V2 activity after imposing a "feedforward" time shift. Then in principle, the top canonical dimension identified from this time-shifted activity could reflect such a feedforward interaction (resembling, for example, DLAG's Across 3 in Fig. 5a, a nominally feedforward latent variable). One could analogously find the top canonical dimension for "feedback-shifted" V1-V2 activity to reveal a feedback interaction (resembling, for example, DLAG's Across 1 in Fig. 5a, a nominally feedback latent variable).

To investigate whether this expectation holds in the V1-V2 recordings, we employed a scheme similar to that of [41] (but modified to better facilitate comparison with DLAG), and studied how projections of V1 and V2 activity onto the top canonical dimension qualitatively change as CCA is fit to V1-V2 activity with different relative time shifts. Specifically, we first took a fixed window of activity in V1, 1240 ms in length, from 20 ms to 1260 ms after stimulus onset. We counted spikes within this window in 20 ms nonoverlapping time bins. For V2, we considered three different (overlapping) time windows, each 1240 ms in length: from 0 ms to 1240 ms after stimulus onset, from 20 ms to 1260 ms after stimulus onset, and from 40 ms to 1280 ms after stimulus onset. In each of these windows, we counted spikes in 20 ms nonoverlapping time bins. We then fit a separate CCA model between the fixed window of activity in V1 and each of the three windows of activity in V2. Then for each fitted model, we projected V1 and V2 neural activity onto the top canonical pair of dimensions.

The projected time courses show no appreciable differences across the three time-shifted model fits. Left: A CCA model was fit to time-shifted activity, in which V2 activity was shifted to lead V1 activity by 20 ms (-20 ms delay). Center: A CCA model was fit to simultaneously observed V1 and V2 activity. Right: A CCA model was fit to time-shifted activity, in which V2 activity was shifted to lag V1 activity by 20 ms (+20 ms delay). Top row / blue box: V1. Bottom row / red box: V2. Each black trace corresponds to one trial; for clarity, only 10 of 400 are shown. All time courses are aligned to stimulus onset. a.u.: arbitrary units.

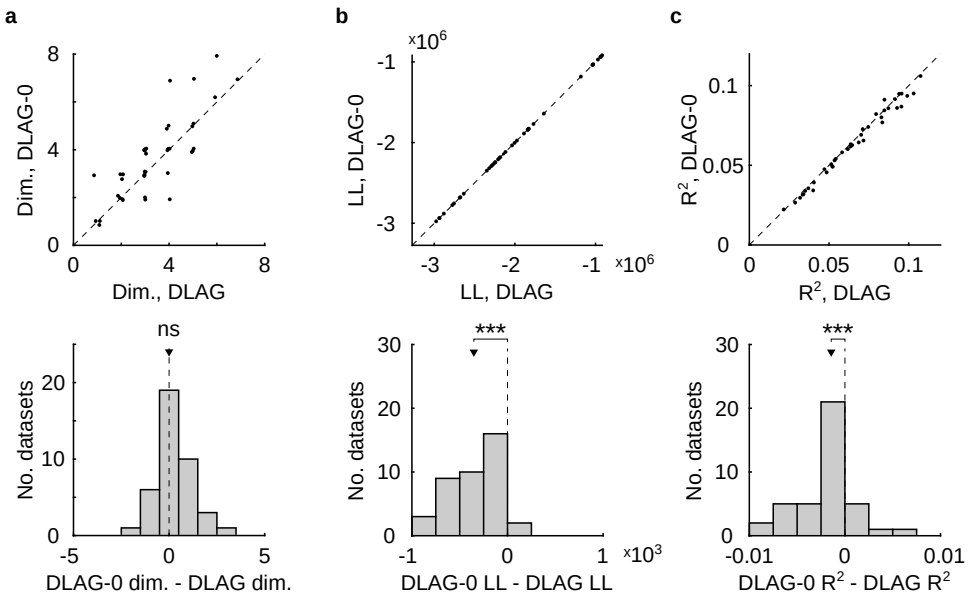
Even though the time courses in each of the three cases look similar, do they reflect signal flow in different directions? To address this question, we estimated a time delay for each pair of fitted canonical dimensions using the same procedure as in Supplementary Fig. 13: we identified the time delay at which projections of V1 activity and projections of V2 activity had maximum cross-correlation. The cross-correlation function between V1 and V2 projections was computed

with 1 ms resolution, from -40 ms (V2 leads V1) to +40 ms (V1 leads V2). In detail, we first took a fixed window of activity in V1, 1200 ms in length, from 40 ms to 1240 ms after stimulus onset. For each trial, we counted spikes within this window in 20 ms nonoverlapping time bins, and projected this activity onto each canonical dimension in V1. For V2, we employed a sliding window of length 1200 ms, which we advanced in 1 ms increments, from 0 ms to 1280 ms after stimulus onset. At each increment, we counted spikes within the window in 20 ms nonoverlapping time bins, and projected this activity (on each trial) onto each canonical dimension in V2. For each canonical pair, we computed the Pearson correlation between the projected V1 activity and the projected V2 activity. This correlation value gave one element of a cross-correlation function: repeating this procedure at each increment of the sliding window in V2 produced a cross-correlation function from -40 ms to +40 ms. We then identified the time delay at which the cross-correlation function for each canonical pair was maximum. Vertical arrows point in the direction of signal flow implied by this time delay.

For the canonical pair fit to V1-V2 observations with a -20 ms time shift (left panel), the identified time delay is indeed negative—but so are the time delays identified in the other two cases. Here, a feedback interaction is dominant, and its cross-correlation (a function of the relative time lag between V1 and V2) decays sufficiently slowly that it remains dominant over a wide range of time lags. Thus the top canonical pair reflects this dominant feedback interaction even when fit to feedforward-shifted V1-V2 activity (right panel). This phenomenon demonstrates the challenge of using a static method like CCA, as we have done here (see also [41]), to disentangle concurrent, bidirectional interactions across areas. We note that, in contrast to the results demonstrated here, [41] found bidirectional (though not concurrent) signals because a much smaller analysis time window was used (80 ms), which enabled the characterization of feedforward- and feedback-dominated trial periods. The concepts demonstrated here still apply within each of those trial periods.

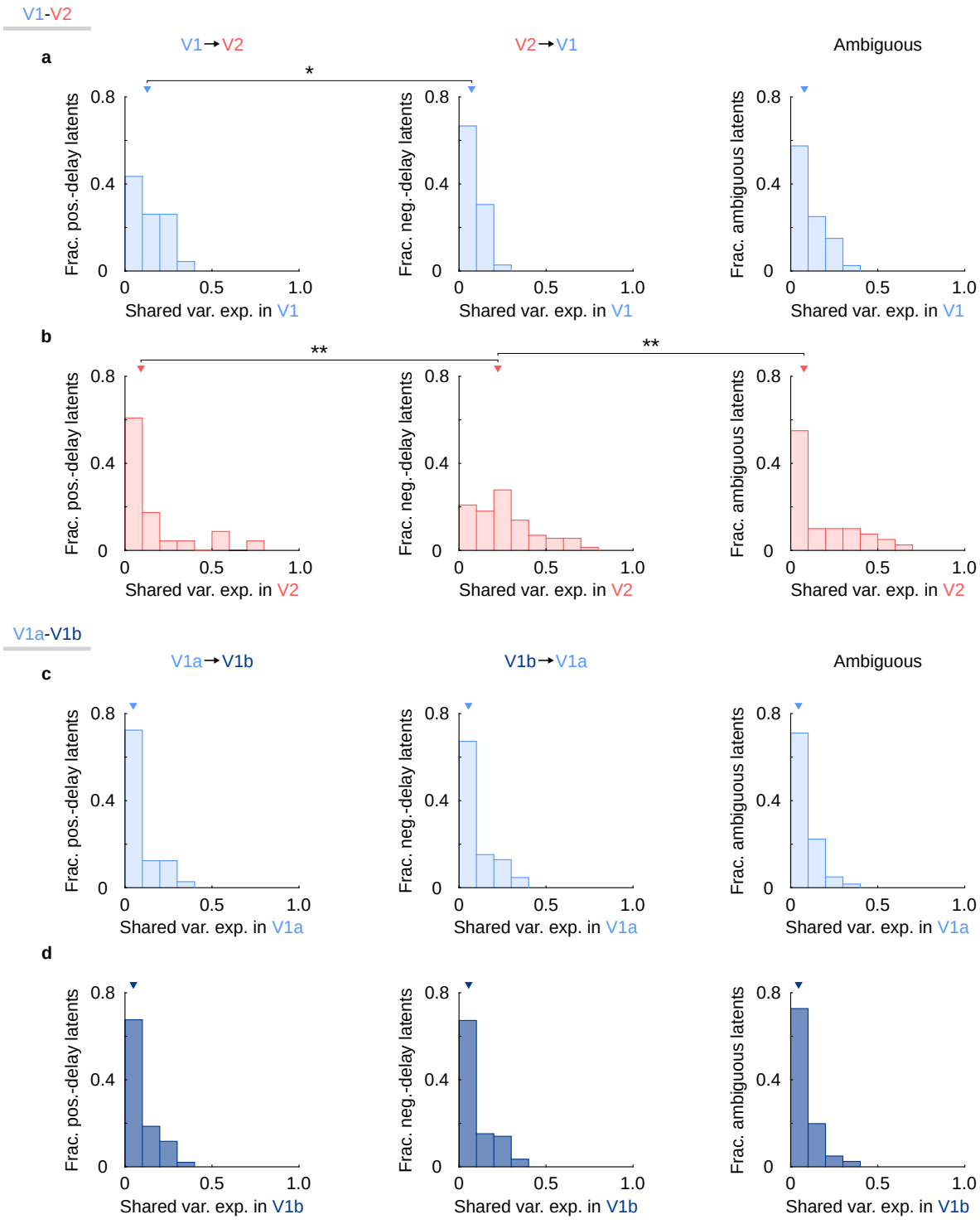
(**b**) Canonical variables fit to simultaneously observed activity indicate only predominant feedback (V2 to V1) activity or zero-lag activity. We again considered the CCA model fit to simultaneously observed activity (panel (a), center), and sought to assess the direction of signal flow associated with all significant canonical pairs selected via cross-validation (see Supplementary Fig. 12). We estimated a time delay for each canonical pair using the same procedure as described in Supplementary Fig. 13 and in panel (a). Canonical variables are paired vertically, and ordered from left to right according to descending canonical correlation value. All other conventions are the same as in (a).

The first canonical pair is associated with a negative (V2 to V1) delay, similar to DLAG's Across 1 and Across 2 in Fig. 5a. But notably, the remaining canonical pairs are associated with time delays at or near 0 ms, whereas DLAG identified a similarly periodic signal with a time delay of +5 ms (Across 3 in Fig. 5a). The qualitative discrepancy between CCA and DLAG could be due to two possible sources: (1) Given the same data, CCA has less statistical power than DLAG, and (2) The mathematical definition of CCA limits its ability to disentangle concurrent signals, irrespective of the amount of available data (as illustrated in Supplementary Fig. 13).



Supplementary Figure 15. V1-V2 interactions are better described by DLAG models with time delays than without time delays. To demonstrate the benefit of including time delays in the statistical model, we re-applied DLAG to the V1-V2 datasets presented in Fig. 6, but forced all time delay parameters to be zero throughout model selection and fitting. We abbreviate these constrained models as 'DLAG-0' from here on and in all figure panels. (**a**) Comparison of DLAG-0 and DLAG across-area dimensionality estimates. For each of the 40 V1-V2 datasets, we identified the number of within- and across-area latent variables for DLAG-0 models using the same two-stage model selection procedure as for the DLAG models (see Methods). Hence estimates for DLAG-0 and DLAG dimensionalities were based on the same first-stage factor analysis (FA) estimates of dimensionality. Top: Estimated DLAG-0 across-area dimensionality ('dim.') versus estimated DLAG across-area dimensionality. Each data point represents one V1-V2 dataset. Data points are integer-valued, but randomly jittered to show points that overlap. DLAG-0 and DLAG estimates of across-area dimensionality are highly correlated (Pearson correlation coefficient, $r = 0.81$). Whether or not the ability to fit time delays leads to higher or lower estimates of across-area dimensionality depends on the idiosyncrasies of the neural activity being analyzed. Greater model flexibility provided by time delays could lead to fewer identified dimensions (cf. [45]). However, the ability to capture time-delayed interactions could also lead to the discovery of additional dimensions that contain significant (time-lagged) cross-area correlations—correlations that would have gone otherwise undetected by a method that could not account for time delays. Bottom: Distribution of the differences between DLAG-0 and DLAG across-area dimensionality estimates on each dataset. 'ns': across-area dimensionality estimates are not significantly different across datasets (one-sided paired sign test: $p = 0.0946$; black triangle indicates the median difference across datasets: 0). (**b**)–(**c**) DLAG outperforms DLAG-0 according to multiple metrics. On each dataset, we compared the optimal DLAG-0 model to the optimal DLAG model (each selected through cross-validation) via two performance metrics: cross-validated data log-likelihood (LL; b) and cross-validated leave-group-out R^2 (c). See Methods for details. Top panels: DLAG-0 performance versus DLAG performance. Each data point represents one V1-V2 dataset. Bottom panels: Distribution of differences between DLAG-0 and DLAG performance on each dataset. DLAG significantly outperforms DLAG-0 across datasets (black triangles indicate the median difference across datasets; '***': one-sided paired sign test; $p < 0.001$). DLAG-0 does

outperform DLAG on some datasets (2 of 40 datasets according to LL; 7 of 40 datasets according to leave-group-out R^2), not inconsistent with the results presented in Fig. 6c, in which many "ambiguous" time delays were identified, whose magnitudes did not significantly deviate from zero (see also Methods). Preferably, one would assess the significance of time delay estimates on a case-by-case basis, as we have done throughout this work.



Supplementary Figure 16. The strongest across-area interactions in V1 are nominally feedforward (V1 to V2), while the strongest across-area interactions in V2 are nominally feedback (V2 to V1). (**a**) Normalized distributions of the fraction of shared variance explained in V1 ('Shared var. exp. in V1') by individual across-area latent variables across all 40 datasets. Left: All across-area latent variables with a significant positive delay (V1 to V2). 'Frac. pos.-delay latents': Fraction of positive-delay latent variables. Center: All across-area latent variables with a

significant negative delay (V2 to V1). 'Frac. neg.-delay latents': Fraction of negative-delay latent variables. Right: All across-area latent variables with an ambiguous delay (not significantly positive or negative). '*': Individual positive-delay latent variables explained more shared variance in V1 than individual negative-delay latent variables (one-sided Wilcoxon rank sum test, $p = 0.042$). (**b**) Normalized distributions of the fraction of shared variance explained in V2 ('Shared var. exp. in V2') by individual across-area latent variables across all 40 datasets. Same conventions as in (a). '**': Individual negative-delay latent variables explained more shared variance in V2 than individual positive-delay latent variables and individual latent variables with ambiguous delays (one-sided Wilcoxon rank sum test, $p < 0.01$). (**c**) Normalized distributions of the fraction of shared variance explained in V1a ('Shared var. exp. in V1a') by individual across-population latent variables across all 40 datasets. Left: All across-population latent variables with a significant positive delay (V1a to V1b). Center: All across-population latent variables with a significant negative delay (V1b to V1a). Right: All across-population latent variables with an ambiguous delay (not significantly positive or negative). No type of latent variable explained more or less shared variance in V1a than any other type of latent variable (two-sided Wilcoxon rank sum test, $p > 0.05$ in all cases). (**d**) Normalized distributions of the fraction of shared variance explained in V1b ('Shared var. exp. in V1b') by individual across-population latent variables across all 40 datasets. Same conventions as in (c). No type of latent variable explained more or less shared variance in V1b than any other type of latent variable (two-sided Wilcoxon rank sum test, $p > 0.05$ in all cases).

Supplementary Note

Fitting the DLAG model

Equations (1)–(8) provide a full definition of the DLAG model. In this section, we describe how DLAG model parameters are fit using exact Expectation Maximization (EM), where the parameters are

$$\theta = \left\{C,\ \mathbf{d},\ R,\ \{D_j\}_{j=1}^{p^a},\ \{\tau_j^a\}_{j=1}^{p^a},\ \{\tau_{1,j}^w\}_{j=1}^{p_1^w},\ \{\tau_{2,j}^w\}_{j=1}^{p_2^w}\right\} \tag{19}$$

Toward that end, we first write the DLAG observation model more compactly as follows. Define the joint activity of neurons in all brain areas by vertically concatenating the observations in each area, $\mathbf{y}_{1,t}$ and $\mathbf{y}_{2,t}$:

$$\mathbf{y}_t = \begin{bmatrix}\mathbf{y}_{1,t}\\ \mathbf{y}_{2,t}\end{bmatrix} \in \mathbb{R}^q \tag{20}$$

where $q = q_1 + q_2$. Next we group together the across- and within-area latent variables for the i^{th} brain area to define $\mathbf{x}_{i,t} = [\mathbf{x}_{i,:,t}^{a\top}\ \mathbf{x}_{i,:,t}^{w\top}]^\top \in \mathbb{R}^{p_i}$, where $p_i = p^a + p_i^w$. We then vertically concatenate the latent variables in each area:

$$\mathbf{x}_t = \begin{bmatrix}\mathbf{x}_{1,t}\\ \mathbf{x}_{2,t}\end{bmatrix} \in \mathbb{R}^p \tag{21}$$

where $p = p_1 + p_2$. We also define the following structured matrices. First define $C_i = [C_i^a\ C_i^w] \in \mathbb{R}^{q_i \times p_i}$ by horizontally concatenating C_i^a and C_i^w. Then, we collect the C_i into a block-diagonal matrix as follows:

$$C = \begin{bmatrix}C_1 & \mathbf{0}\\ \mathbf{0} & C_2\end{bmatrix} \in \mathbb{R}^{q\times p} \tag{22}$$

Similarly, define

$$R = \begin{bmatrix}R_1 & \mathbf{0}\\ \mathbf{0} & R_2\end{bmatrix} \in \mathbb{R}^{q\times q}, \tag{23}$$

$$\mathbf{d} = \begin{bmatrix}\mathbf{d}_1\\ \mathbf{d}_2\end{bmatrix} \in \mathbb{R}^q \tag{24}$$

We can then write the DLAG observation model compactly as follows:

$$\mathbf{y}_t \mid \mathbf{x}_t \sim \mathcal{N}(C\mathbf{x}_t + \mathbf{d}, R) \tag{25}$$

The observation model expressed in equation (25) defines a distribution for neural activity at a single time point, but to properly fit the DLAG model, we must consider the distribution over all time points. Thus we define $\bar{\mathbf{y}} = [\mathbf{y}_1^\top \cdots \mathbf{y}_T^\top]^\top \in \mathbb{R}^{qT}$ and $\bar{\mathbf{x}} = [\mathbf{x}_1^\top \cdots \mathbf{x}_T^\top]^\top \in \mathbb{R}^{pT}$, obtained by vertically concatenating the observed variables $\mathbf{y}_t$ and latent variables $\mathbf{x}_t$, respectively, across all $t = 1, \ldots, T$. Then, we rewrite the state and observation models as follows:

$$\bar{\mathbf{x}} \sim \mathcal{N}(\mathbf{0}, \bar{K}) \tag{26}$$

$$\bar{\mathbf{y}} \mid \bar{\mathbf{x}} \sim \mathcal{N}(\bar{C}\bar{\mathbf{x}} + \bar{\mathbf{d}}, \bar{R}), \tag{27}$$

where $\bar{C} \in \mathbb{R}^{qT\times pT}$ and $\bar{R} \in \mathbb{S}^{qT\times qT}$ are block diagonal matrices comprising T copies of the matrices C and R, respectively. $\bar{\mathbf{d}} \in \mathbb{R}^{qT}$ is constructed by vertically concatenating T copies of $\mathbf{d}$. The elements of $\bar{K} \in \mathbb{R}^{pT\times pT}$ are computed using equations (3)–(8). Then, the joint distribution over observed and latent variables is given by

$$\begin{bmatrix}\bar{\mathbf{x}}\\ \bar{\mathbf{y}}\end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix}\mathbf{0}\\ \bar{\mathbf{d}}\end{bmatrix}, \begin{bmatrix}\bar{K} & \bar{K}\bar{C}^\top\\ \bar{C}\bar{K} & \bar{C}\bar{K}\bar{C}^\top + \bar{R}\end{bmatrix}\right) \tag{28}$$

E-step In the E-step, our goal is to compute the posterior distribution of the latent variables $\bar{\mathbf{x}}$ given the recorded neural activity $\bar{\mathbf{y}}$, $P(\bar{\mathbf{x}}|\bar{\mathbf{y}})$, using the most recent parameter estimates θ. Using basic results of conditioning for jointly Gaussian random variables, we get

$$\bar{\mathbf{x}} \mid \bar{\mathbf{y}} \sim \mathcal{N}\left(\bar{K}\bar{C}^\top\left(\bar{C}\bar{K}\bar{C}^\top + \bar{R}\right)^{-1}\left(\bar{\mathbf{y}} - \bar{\mathbf{d}}\right), \bar{K} - \bar{K}\bar{C}^\top\left(\bar{C}\bar{K}\bar{C}^\top + \bar{R}\right)^{-1}\bar{C}\bar{K}\right) \tag{29}$$

Thus, posterior estimates of latent variables are given by

$$\mathbb{E}\left[\bar{\mathbf{x}}|\bar{\mathbf{y}}\right] = \bar{K}\bar{C}^\top\left(\bar{C}\bar{K}\bar{C}^\top + \bar{R}\right)^{-1}\left(\bar{\mathbf{y}} - \bar{\mathbf{d}}\right) \tag{30}$$

The marginal likelihood of the observed neural activity can be computed as

$$\bar{\mathbf{y}} \sim \mathcal{N}\left(\bar{\mathbf{d}}, \bar{C}\bar{K}\bar{C}^\top + \bar{R}\right) \tag{31}$$

M-step In the M-step, our goal is to maximize $\mathcal{E}(\theta) = \mathbb{E}\left[\log P(\bar{\mathbf{x}}, \bar{\mathbf{y}})|\theta\right]$ with respect to θ, using the latest inference of the latent variables, computed in the E-step. As in [44, 45], we adopt the following notation. Given a vector $\mathbf{v}$,

$$\langle \mathbf{v} \rangle = \mathbb{E}\left[\mathbf{v}|\bar{\mathbf{y}}\right] \tag{32}$$

$$\langle \mathbf{v}\mathbf{v}^\top \rangle = \mathbb{E}\left[\mathbf{v}\mathbf{v}^\top|\bar{\mathbf{y}}\right] \tag{33}$$

The appropriate expectations can be found using equation (29).

Maximizing $\mathcal{E}(\theta)$ with respect to C, $\mathbf{d}$ yields the following closed-form update for the i^{th} brain area:

$$\begin{bmatrix} C_i & \mathbf{d}_i \end{bmatrix} = \left(\sum_{t=1}^{T} \mathbf{y}_{i,t} \cdot \begin{bmatrix} \langle \mathbf{x}_{i,t} \rangle^\top & 1 \end{bmatrix}\right)\left(\sum_{t=1}^{T} \begin{bmatrix} \langle \mathbf{x}_{i,t}\mathbf{x}_{i,t}^\top \rangle & \langle \mathbf{x}_{i,t} \rangle \\ \langle \mathbf{x}_{i,t} \rangle^\top & 1 \end{bmatrix}\right)^{-1} \tag{34}$$

After performing the update for each area separately, we collect all updated values into C and $\mathbf{d}$. Then we update R for both brain areas together, as follows:

$$R = \frac{1}{T}\text{diag}\left\{\sum_{t=1}^{T}\left((\mathbf{y}_t - \mathbf{d})(\mathbf{y}_t - \mathbf{d})^\top - (\mathbf{y}_t - \mathbf{d})\langle \mathbf{x}_t \rangle^\top C^\top - C\langle \mathbf{x}_t \rangle(\mathbf{y}_t - \mathbf{d})^\top + C\langle \mathbf{x}_t\mathbf{x}_t^\top \rangle C^\top\right)\right\} \tag{35}$$

There are no closed-form solutions for the Gaussian process parameter updates, but we can compute gradients and perform gradient ascent. Note that, for this work, we choose not to fit the Gaussian process noise variances, but rather, we set them to small values (10^{-3}), as in [44]. Within-area timescale gradients for the i^{th} brain area and j^{th} within-area latent variable are given by

$$\frac{\partial \mathcal{E}(\theta)}{\partial \tau_{i,j}^w} = \text{tr}\left(\left(\frac{\partial \mathcal{E}(\theta)}{\partial K_{i,j}^w}\right)^\top\left(\frac{\partial K_{i,j}^w}{\partial \tau_{i,j}^w}\right)\right) \tag{36}$$

where

$$\frac{\partial \mathcal{E}(\theta)}{\partial K_{i,j}^w} = -\frac{1}{2}(K_{i,j}^w)^{-1} + \frac{1}{2}\left((K_{i,j}^w)^{-1}\langle \mathbf{x}_{i,j,:}^w \mathbf{x}_{i,j,:}^{w\top} \rangle (K_{i,j}^w)^{-1}\right) \tag{37}$$

and element (t_1, t_2) of $\partial K^w_{i,j}/\partial \tau^w_{i,j}$ is given by

$$\frac{\partial k^w_{i,j}(t_1,t_2)}{\partial \tau^w_{i,j}} = \left(1-(\sigma^w_{i,j})^2\right)\frac{(t_2-t_1)^2}{(\tau^w_{i,j})^3}\exp\left(-\frac{(t_2-t_1)^2}{2(\tau^w_{i,j})^2}\right) \tag{38}$$

To express the across-area timescale and delay parameter gradients, we introduce more compact notation for the variables in equation (6). Let $\mathbf{x}^a_{j,:} = [\mathbf{x}^{a\top}_{1,j,:}\mathbf{x}^{a\top}_{2,j,:}]^\top \in \mathbb{R}^{2T}$ for the j^{th} across-area latent variable, and

$$K^a_j = \begin{bmatrix} K^a_{1,1,j} & K^a_{1,2,j} \\ K^a_{2,1,j} & K^a_{2,2,j} \end{bmatrix} \in \mathbb{S}^{2T\times 2T} \tag{39}$$

Then, across-area timescale gradients are given by

$$\frac{\partial \mathcal{E}(\theta)}{\partial \tau^a_j} = \operatorname{tr}\left(\left(\frac{\partial \mathcal{E}(\theta)}{\partial K^a_j}\right)^\top\left(\frac{\partial K^a_j}{\partial \tau^a_j}\right)\right) \tag{40}$$

where

$$\frac{\partial \mathcal{E}(\theta)}{\partial K^a_j} = -\frac{1}{2}(K^a_j)^{-1} + \frac{1}{2}\left((K^a_j)^{-1}\langle \mathbf{x}^a_{j,:}\mathbf{x}^{a\top}_{j,:}\rangle(K^a_j)^{-1}\right) \tag{41}$$

and each element of $\partial K^a_j/\partial \tau^a_j$ is given by

$$\frac{\partial k^a_{i_1,i_2,j}(t_1,t_2)}{\partial \tau^a_j} = \left(1-(\sigma^a_j)^2\right)\frac{(\Delta t)^2}{(\tau^a_j)^3}\exp\left(-\frac{(\Delta t)^2}{2(\tau^a_j)^2}\right) \tag{42}$$

where Δt is defined as in equation (8). To optimize the timescales while respecting non-negativity constraints, we perform a change of variables, and then perform unconstrained gradient ascent with respect to $\log \tau^w_{i,j}$ or $\log \tau^a_j$.

Next, delay gradients for brain area i and across-area latent variable j are given by

$$\frac{\partial \mathcal{E}(\theta)}{\partial D_{i,j}} = \operatorname{tr}\left(\left(\frac{\partial \mathcal{E}(\theta)}{\partial K^a_j}\right)^\top\left(\frac{\partial K^a_j}{\partial D_{i,j}}\right)\right) \tag{43}$$

where $\frac{\partial \mathcal{E}(\theta)}{\partial K^a_j}$ is defined as in equation (41), and each element of $\partial K^a_j/\partial D_{i,j}$ is given by

$$\frac{\partial k^a_{i_1,i_2,j}(t_1,t_2)}{\partial D_{i,j}} = \left(1-(\sigma^a_j)^2\right)\frac{\Delta t}{(\tau^a_j)^2}\exp\left(-\frac{(\Delta t)^2}{2(\tau^a_j)^2}\right)\frac{\partial(\Delta t)}{\partial D_{i,j}} \tag{44}$$

$$\frac{\partial(\Delta t)}{D_{i,j}} = \begin{cases} 1 & \text{if} \quad i = i_2 \\ -1 & \text{if} \quad i = i_1 \end{cases} \tag{45}$$

where Δt, i_1, and i_2 are defined as in equation (8). In practice, we fix all delay parameters for area 1 at 0 to ensure identifiability. As with the timescales, one might wish to constrain the delays within some physically realistic range, such as the length of an experimental trial, so that $-D_{\max} \le D_{i,j} \le D_{\max}$. Toward that end, we make the change of variables $D_{i,j} = D_{\max}\frac{1-e^{-D^*_{i,j}}}{1+e^{-D^*_{i,j}}}$ and perform unconstrained gradient ascent with respect to $D^*_{i,j}$. Here we chose $D_{\max}$ to be half the length of a trial. No delays came close to these constraints in our results (Fig. 6, Supplementary Fig. 9).

Finally, note that all of these EM updates are derived for a single sequence, or trial. It is straightforward to extend these equations to N independent trials (each with a potentially different number of time steps, T) by maximizing $\frac{\partial}{\partial\theta}\left[\sum_{n=1}^{N}\mathcal{E}_n(\theta)\right]$, where trial is indexed by $n = 1, \ldots, N$.

Parameter initialization To initialize the DLAG observation model parameters to reasonable values prior to fitting with the EM algorithm, we first fit a probabilistic canonical correlation analysis (pCCA) [42] model to the neural activity, with the same number of across-area latent variables as the desired DLAG model (see next section). pCCA is defined by the following state and observation models:

$$\mathbf{x}_t^a \sim \mathcal{N}(\mathbf{0}, I) \tag{46}$$

$$\mathbf{y}_{i,t} \mid \mathbf{x}_t^a \sim \mathcal{N}(C_i^a \mathbf{x}_t^a + \mathbf{d}_i, R_i) \tag{47}$$

where $C_i^a \in \mathbb{R}^{q_i \times p^a}$ maps the p^a-dimensional across-area latent variables $\mathbf{x}_t^a \in \mathbb{R}^{p^a}$ to the neural activity of area i, $\mathbf{d}_i \in \mathbb{R}^{q_i}$ is a mean parameter, and $R_i \in \mathbb{S}^{q_i \times q_i}$ is the observation noise covariance matrix. R_i is not constrained to be diagonal. The fitted values for C_i^a and $\mathbf{d}_i$ are used as initial values for their DLAG analogues. We take only the diagonal elements of R_i to initialize its DLAG analogue.

pCCA does not incorporate within-area latent variables. Therefore, we initialized each DLAG within-area loading matrix C_i^w so that its columns spanned a subspace uncorrelated with that spanned by the columns of C_i^a, returned by pCCA. Such a subspace can be computed as follows. Let $\Sigma_i \in \mathbb{S}^{q_i \times q_i}$ be the sample covariance matrix of activity in area i. Then define $M_i = C_i^{a\top}\Sigma_i \in \mathbb{R}^{p^a \times q_i}$. The singular value decomposition of M_i is given by $M_i = U_i S_i V_i^\top$, where $U_i \in \mathbb{R}^{p^a \times p^a}$, $S_i \in \mathbb{R}^{p^a \times q_i}$, and $V_i \in \mathbb{R}^{q_i \times q_i}$. The first p^a columns of V_i span the same across-area subspace spanned by the columns of C_i^a. The remaining $q_i - p^a$ columns form an orthonormal basis for the subspace uncorrelated with this across-area subspace. We initialized C_i^w with the first p_i^w of these uncorrelated basis vectors. Finally, we initialized all delays to zero, and all within- and across-area Gaussian process timescales to the same value, equal to twice the sampling period or spike count bin width of the neural activity.

Supplementary Discussion

Statistical tradeoffs between within- and across-area latent variables

Throughout this work, we have described how DLAG decomposes observed neural activity into a linear combination of within- and across-area latent variables. Equivalently, DLAG partitions each area's population space into distinct within- and across-area subspaces, which represent characteristic ways in which the neurons covary (Fig. 2). Here we investigate more deeply why the within-area latent variables are a necessary model component, even if across-area activity is of primary scientific interest. Toward that end, we will consider an alternative interpretational perspective: namely, that DLAG performs a low-rank decomposition of the covariance matrix of a time series. This alternative perspective also illuminates a general statistical phenomenon—not specific to DLAG—that any multi-area time series method must consider.

DLAG performs a low-rank covariance decomposition

Let us first express the DLAG model not only for a single time point, as in equation (25), but for all time points in a sequence. In particular, we will collect observed and latent variables in a manner that highlights group structure (i.e., organized differently than in equations (26) and (27)). We define $\widetilde{\mathbf{y}}_1 = [\mathbf{y}_{1,1}^\top \cdots \mathbf{y}_{1,T}^\top]^\top \in \mathbb{R}^{q_1 T}$ and $\widetilde{\mathbf{y}}_2 = [\mathbf{y}_{2,1}^\top \cdots \mathbf{y}_{2,T}^\top]^\top \in \mathbb{R}^{q_2 T}$, obtained by vertically concatenating the observed neural activity $\mathbf{y}_{1,t}$ and $\mathbf{y}_{2,t}$ in areas 1 and 2, respectively, across all times $t = 1, \ldots, T$. We collect the across- and within-area latent variables for each area similarly. Let $\widetilde{\mathbf{x}}_1^a = [\mathbf{x}_{1,:,1}^{a\top} \cdots \mathbf{x}_{1,:,T}^{a\top}]^\top \in \mathbb{R}^{p^a T}$, $\widetilde{\mathbf{x}}_1^w = [\mathbf{x}_{1,:,1}^{w\top} \cdots \mathbf{x}_{1,:,T}^{w\top}]^\top \in \mathbb{R}^{p_1^w T}$, $\widetilde{\mathbf{x}}_2^a = [\mathbf{x}_{2,:,1}^{a\top} \cdots \mathbf{x}_{2,:,T}^{a\top}]^\top \in \mathbb{R}^{p^a T}$, and $\widetilde{\mathbf{x}}_2^w = [\mathbf{x}_{2,:,1}^{w\top} \cdots \mathbf{x}_{2,:,T}^{w\top}]^\top \in \mathbb{R}^{p_2^w T}$.

Then, we rewrite the state and observation models as follows:

$$\begin{bmatrix} \widetilde{\mathbf{x}}_1^a \\ \widetilde{\mathbf{x}}_1^w \\ \widetilde{\mathbf{x}}_2^a \\ \widetilde{\mathbf{x}}_2^w \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \widetilde{K}_{1,1}^a & \mathbf{0} & \widetilde{K}_{1,2}^a & \mathbf{0} \\ \mathbf{0} & \widetilde{K}_1^w & \mathbf{0} & \mathbf{0} \\ \widetilde{K}_{2,1}^a & \mathbf{0} & \widetilde{K}_{2,2}^a & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \widetilde{K}_2^w \end{bmatrix} \right) \tag{48}$$

$$\begin{bmatrix} \widetilde{\mathbf{y}}_1 \\ \widetilde{\mathbf{y}}_2 \end{bmatrix} \mid \begin{bmatrix} \widetilde{\mathbf{x}}_1^a \\ \widetilde{\mathbf{x}}_1^w \\ \widetilde{\mathbf{x}}_2^a \\ \widetilde{\mathbf{x}}_2^w \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \widetilde{C}_1^a & \widetilde{C}_1^w & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \widetilde{C}_2^a & \widetilde{C}_2^w \end{bmatrix} \begin{bmatrix} \widetilde{\mathbf{x}}_1^a \\ \widetilde{\mathbf{x}}_1^w \\ \widetilde{\mathbf{x}}_2^a \\ \widetilde{\mathbf{x}}_2^w \end{bmatrix} + \begin{bmatrix} \widetilde{\mathbf{d}}_1 \\ \widetilde{\mathbf{d}}_2 \end{bmatrix}, \begin{bmatrix} \widetilde{R}_1 & \mathbf{0} \\ \mathbf{0} & \widetilde{R}_2 \end{bmatrix} \right) \tag{49}$$

where $\widetilde{C}_1^a \in \mathbb{R}^{q_1 T \times p^a T}$, $\widetilde{C}_1^w \in \mathbb{R}^{q_1 T \times p_1^w T}$, $\widetilde{C}_2^a \in \mathbb{R}^{q_2 T \times p^a T}$, $\widetilde{C}_2^w \in \mathbb{R}^{q_2 T \times p_2^w T}$, $\widetilde{R}_1 \in \mathbb{S}^{q_1 T \times q_1 T}$, and $\widetilde{R}_2 \in \mathbb{S}^{q_2 T \times q_2 T}$ are all block diagonal matrices comprising T copies of the loading matrices C_1^a, C_1^w, C_2^a, and C_2^w, and observation noise covariance matrices R_1 and R_2, respectively. $\widetilde{\mathbf{d}}_1 \in \mathbb{R}^{q_1 T}$ and $\widetilde{\mathbf{d}}_2 \in \mathbb{R}^{q_2 T}$ are constructed by vertically concatenating T copies of mean parameters $\mathbf{d}_1$ and $\mathbf{d}_2$, respectively. Note that equations (48) and (49) above are equivalent to equations (26) and (27), but with variables rearranged.

Each within-area covariance matrix $\widetilde{K}_i^w \in \mathbb{S}^{p_i^w T \times p_i^w T}$, for area $i = 1, 2$ has the following block structure:

$$\widetilde{K}_i^w = \begin{bmatrix} \widetilde{K}_i^w(1,1) & \cdots & \widetilde{K}_i^w(1,T) \\ \vdots & \ddots & \vdots \\ \widetilde{K}_i^w(T,1) & \cdots & \widetilde{K}_i^w(T,T) \end{bmatrix} \tag{50}$$

where each block $\widetilde{K}_i^w(t_1, t_2) = \mathrm{diag}(k_{i,1}^w(t_1, t_2), \ldots, k_{i,p_i^w}^w(t_1, t_2)) \in \mathbb{S}^{p_i^w \times p_i^w}, t_1, t_2 \in \{1, \ldots, T\}$ is a diagonal matrix whose elements are computed according to the covariance function defined in equations (4) and (5).

Each across-area auto- or cross-covariance matrix $\widetilde{K}_{i_1,i_2}^a \in \mathbb{R}^{p^a T \times p^a T}$, for areas $i_1, i_2 \in \{1, 2\}$ has analogous structure:

$$\widetilde{K}_{i_1,i_2}^a = \begin{bmatrix} \widetilde{K}_{i_1,i_2}^a(1,1) & \cdots & \widetilde{K}_{i_1,i_2}^a(1,T) \\ \vdots & \ddots & \vdots \\ \widetilde{K}_{i_1,i_2}^a(T,1) & \cdots & \widetilde{K}_{i_1,i_2}^a(T,T) \end{bmatrix} \tag{51}$$

where each block $\widetilde{K}_{i_1,i_2}^a(t_1, t_2) = \mathrm{diag}(k_{i_1,i_2,1}^a(t_1, t_2), \ldots, k_{i_1,i_2,p^a}^a(t_1, t_2)) \in \mathbb{S}^{p^a \times p^a}, t_1, t_2 \in \{1, \ldots, T\}$ is a diagonal matrix whose elements are computed according to the covariance function defined in equations (7) and (8). Note that the cross-covariance matrices are transposes of one another, i.e., $\widetilde{K}_{i_1,i_2}^a = \widetilde{K}_{i_2,i_1}^{a\top}$.

Upon inspection of equation (48), the statistical dependency between latent variables becomes clear. However, the statistical dependency between observed neural activity in each area, $\widetilde{\mathbf{y}}_1$ and $\widetilde{\mathbf{y}}_2$, is not obvious, since the structure of equation (49) suggests that they might be decoupled. The relationship between observed areas becomes clear when we consider their joint distribution, after marginalizing out the latent variables:

$$\begin{bmatrix} \widetilde{\mathbf{y}}_1 \\ \widetilde{\mathbf{y}}_2 \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \widetilde{\mathbf{d}}_1 \\ \widetilde{\mathbf{d}}_2 \end{bmatrix}, \widetilde{\Sigma} \right) \tag{52}$$

where

$$\widetilde{\Sigma} = \begin{bmatrix} \widetilde{C}_1^a \widetilde{K}_{1,1}^a \widetilde{C}_1^{a\top} + \widetilde{C}_1^w \widetilde{K}_1^w \widetilde{C}_1^{w\top} + \widetilde{R}_1 & \widetilde{C}_1^a \widetilde{K}_{1,2}^a \widetilde{C}_2^{a\top} \\ \widetilde{C}_2^a \widetilde{K}_{2,1}^a \widetilde{C}_1^{a\top} & \widetilde{C}_2^a \widetilde{K}_{2,2}^a \widetilde{C}_2^{a\top} + \widetilde{C}_2^w \widetilde{K}_2^w \widetilde{C}_2^{w\top} + \widetilde{R}_2 \end{bmatrix} \tag{53}$$

Equation (53) makes explicit the alternative interpretational perspective of DLAG: DLAG performs a low-rank decomposition of the covariance matrix $\widetilde{\Sigma}$. This decomposition is illustrated graphically in Supplementary Fig. 17a. For simplicity, we illustrate a covariance matrix for areas with three neurons each, over two time points. The shading of blocks of the covariance matrix illustrate which type of DLAG parameter is responsible for explaining that particular portion of covariance (magenta: across-area; blue/red: within-area; gray: independent single-neuron variability). Regions of overlap (i.e., where both blue/magenta or red/magenta shading are present) illustrate portions of covariance that both within- and across-area variables are responsible for explaining. Any regions of white indicate that no model parameters explain that portion of covariance.

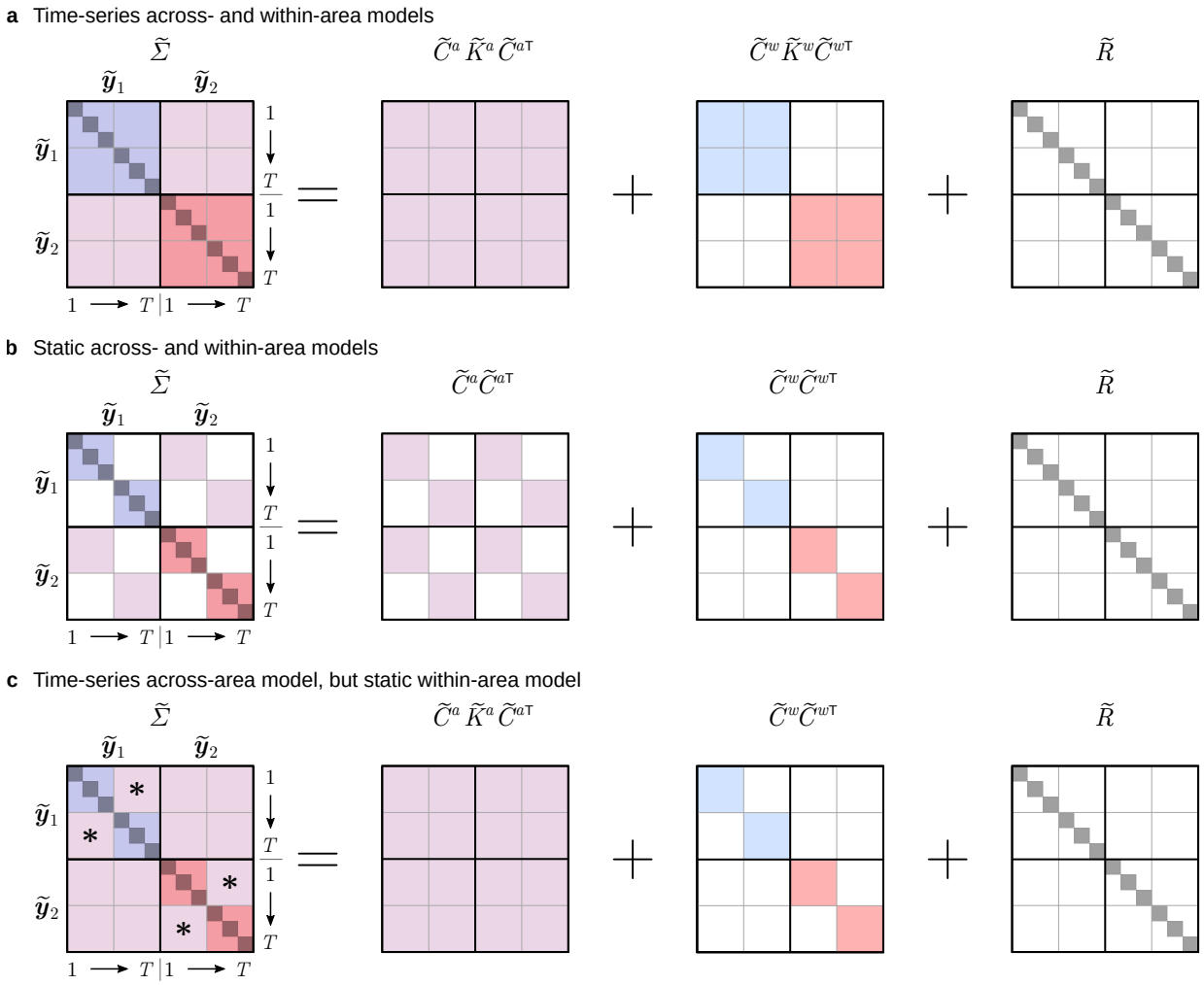
The across-area parameters (note the fully magenta-shaded across-area covariance component in Supplementary Fig. 17a) serve to explain covariance among all neurons, in both areas. Within-area parameters (blue and red shading, for areas 1 and 2, respectively) serve to explain covariance among neurons within each area, but not across areas (note the white across-area blocks for

the within-area covariance component). Importantly, the only parameters in the DLAG model capable of explaining covariance across areas are the across-area parameters (only magenta shading is present in the across-area blocks of $\widetilde{\Sigma}$). And interestingly, within-area components fully overlap across-area components in the within-area blocks of $\widetilde{\Sigma}$, suggesting a potential redundancy. However, as we will discuss below, the overall structure of the decomposition shown in Supplementary Fig. 17a is critical to the interpretation of across-area variables—that they isolate neural interactions *across* areas (and minimally reflect purely within-area interactions).

A time series within-area model must accompany a time series across-area model

To build further intuition, let us consider the scenario where within- and across-area covariances are modeled statically—without considering the flow of time (Supplementary Fig. 17b). Static covariance decompositions result, for example, from the probabilistic canonical correlation analysis (pCCA) model [42], which includes static across-area latent variables and no within-area latent variables (within-area covariance is instead captured using full observation noise covariance matrices, R_1 and R_2). The covariance matrix $\widetilde{\Sigma}$ still decomposes into across- and within-area components; however, covariances at non-zero time lags (i.e., the covariance between neural activity at a time point t_1 and a different time point $t_2 \neq t_1$, indicated by the white-shaded blocks of $\widetilde{\Sigma}$ in Supplementary Fig. 17b) are all zero, by definition. Just like the DLAG case (Supplementary Fig. 17a), only the across-area parameters can explain across-area covariance, and within-area components fully overlap across-area components in the within-area blocks of $\widetilde{\Sigma}$ (to understand why this covariance structure is important, see case below). Across-area activity is successfully isolated by across-area variables.

The problematic case arises when we use a time series model to describe across-area interactions, but use a static model to describe within-area interactions (Supplementary Fig. 17c). For example, what if we proposed a version of DLAG that simply adopted the same observation model as pCCA (i.e., full observation noise covariance matrices, R_1 and R_2) to model within-area interactions? In this case, although the within-area model components do explain covariance among neurons within each area, they fail to capture any within-area covariance across time points, by definition. This shortcoming forces the across-area variables to explain within-area covariance across time points. Visually, all within-area blocks of the covariance matrix $\widetilde{\Sigma}$ representing relationships across time points have solely magenta shading (these problematic blocks are highlighted by the '*' symbols in Supplementary Fig. 17c). In contrast, the true DLAG model and fully static models avoid this pitfall. These successful models (Supplementary Fig. 17a,b) do not have any blocks of $\widetilde{\Sigma}$ for which across-area parameters are solely responsible for explaining within-area covariance. This statistical phenomenon applies to any multi-area time series method, and is not specific to DLAG [32, 62].



Supplementary Figure 17. Full-sequence (trial) covariance matrix decompositions. For simplicity, in (a)-(c), we illustrate a covariance matrix for areas with three neurons each, over two time points. From left to right, panels represent the overall covariance matrix, its across-area component, its within-area component, and a component representing variance independent to each neuron. Across-area parameters (magenta shading) are solely responsible for explaining across-area covariance over time (i.e., there is no overlap of magenta with blue, red, or gray in the across-area off-diagonal blocks of the overall covariance matrix, on the left). (**a**) DLAG decomposes the covariance of a full sequence (trial) into low-rank components. Covariance among neurons within an area that cannot be explained by across-area covariance is captured by within-area parameters (area A: blue; area B: red). (**b**) Models such as probabilistic canonical correlation analysis (pCCA), for example, similarly decompose the overall covariance matrix into across- and within-area components, but make no attempt to model covariance across time points, either across or within areas (indicated by blocks with white shading). (**c**) If one is using a time series across-area model, then in the absence of a time series within-area model, across-area parameters are forced to explain within-area covariance over time. This problem is illustrated by the within-area blocks of the overall covariance matrix that have only magenta shading (indicated by the '*' symbols).

Supplementary References

[72] Williamson, R. C. *et al.* Scaling properties of dimensionality reduction for neural populations and network models. *PLOS Computational Biology* **12**, e1005141 (2016).

[73] Freeman, J., Ziemba, C. M., Heeger, D. J., Simoncelli, E. P. & Movshon, J. A. A functional and perceptual signature of the second visual area in primates. *Nature Neuroscience* **16**, 974–981 (2013).

[74] Okazawa, G., Tajima, S. & Komatsu, H. Gradual development of visual texture-selective properties between macaque areas V2 and V4. *Cerebral Cortex* **27**, 4867–4880 (2017).

[75] Jasper, A. I., Tanabe, S. & Kohn, A. Predicting perceptual decisions using visual cortical population responses and choice history. *Journal of Neuroscience* **39**, 6714–6727 (2019).

[76] Portilla, J. & Simoncelli, E. P. A parametric texture model based on joint statistics of complex wavelet coefficients. *International Journal of Computer Vision* **40**, 49–70 (2000).