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

nature computational science

Disentangling the flow of signals between populations of neurons

Evren Gokcen[1], Anna I. Jasper[2], João D. Semedo[1], Amin Zandvakili[2], Adam Kohn[2,3,4,7], Christian K. Machens[5,7] and Byron M. Yu[1,6,7] ✉

Technological advances now allow us to record from large populations of neurons across multiple brain areas. These recordings may illuminate how communication between areas contributes to brain function, yet a substantial barrier remains: how do we disentangle the concurrent, bidirectional flow of signals between populations of neurons? We propose here a dimensionality reduction framework, delayed latents across groups (DLAG), that disentangles signals relayed in each direction, identifies how these signals are represented by each population and characterizes how they evolve within and across trials. We demonstrate that DLAG performs well on synthetic datasets similar in scale to current neurophysiological recordings. Then we study simultaneously recorded populations in primate visual areas V1 and V2, where DLAG reveals signatures of bidirectional yet selective communication. Our framework lays a foundation for dissecting the intricate flow of signals across populations of neurons, and how this signalling contributes to cortical computation.

Simultaneous recordings from large populations of neurons across multiple brain areas are growing in availability[1–4]. These recordings present opportunities to illuminate how inter-areal communication enables brain function[5], but they also present substantial conceptual and statistical challenges. Brain areas involved in sensory[6–9], cognitive[10] and motor functions[11] are often reciprocally connected: signals are relayed not only from one area to the next, but bidirectionally, and probably concurrently. The raw recordings, however, provide only a tangled view of this concurrent communication (Fig. 1, top): individual neurons simultaneously reflect an area's inputs, outputs and ongoing internal computations[12].

Determining the flow of signals between brain areas is therefore a non-trivial task. To dissect the direction of signal flow, one can leverage the fact that inter-areal communication is not instantaneous. The physiological properties of axons and synapses introduce delays in signal transmission. These delays provide a working definition of signal flow: the appearance of a signal first in area A, and later in area B, is consistent with signal flow from A to B (though this apparent flow could be due to common input from a third area; see Discussion).

Adopting this conception, several inter-areal studies have compared the timing of the onset of neural responses[13–15] or of the emergence of selectivity attributable to top-down processes[16–20] across areas following the presentation of a stimulus. Other studies, leveraging simultaneous recordings, have measured temporal delays between two areas through pairwise spiking correlations[21–26] and information-theoretic measures[27]. Similarly, inter-areal phase delays of local field potentials have been measured[28–31]. These timing-based approaches have advanced our understanding of how signals propagate across brain areas. However, because these approaches focus largely on pairs of neurons or aggregate measures of neural activity, much remains unknown about how neuronal populations coordinate their activity to accomplish inter-areal signalling.

To characterize inter-areal signal flow at the level of neuronal populations is a challenging high-dimensional problem. Dimensionality reduction techniques capable of identifying low-dimensional latent variables that describe activity shared by two or more recorded areas are thus increasingly used[32–34]. These techniques have driven new proposals for population-level mechanisms of gating between motor cortex output and muscle movement[35], selective communication between cortical areas[36,37], enhanced communication of stimulus information with attention[38] and the robustness of local computations to perturbations upstream[39,40].

The relationship between the correlated activity across areas identified in these studies and the flow of inter-areal signals, however, remains unclear. Specifically, does the correlated activity across areas reflect the flow of activity from area A to B, from B to A, or in both directions concurrently (Fig. 1, bottom left)? If communication were to occur in one direction at a time, then existing dimensionality reduction methods could, in principle, identify the direction of population-level signal flow. If two areas were to communicate in both directions concurrently, however, then existing methods would only identify the dominant direction of signal flow[41].

We therefore propose delayed latents across groups (DLAG; Fig. 1, bottom right), a dimensionality reduction framework that disentangles signals relayed in each direction, identifies how these signals are represented by each population and characterizes how they evolve over time within and across trials. We first demonstrate that DLAG performs well on synthetic datasets similar in scale to current neurophysiological data. Then we study simultaneously recorded populations in primate visual areas V1 and V2, where DLAG reveals that V1–V2 interactions are selective and bidirectional. DLAG unlocks new opportunities to investigate the bidirectional flow of signals between populations of neurons and how inter-areal communication contributes to brain function.

[1]Department of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, PA, USA. [2]Dominick Purpura Department of Neuroscience, Albert Einstein College of Medicine, New York, NY, USA. [3]Department of Ophthalmology and Visual Sciences, Albert Einstein College of Medicine, New York, NY, USA. [4]Department of Systems and Computational Biology, Albert Einstein College of Medicine, New York, NY, USA. [5]Champalimaud Neuroscience Programme, Champalimaud Foundation, Lisbon, Portugal. [6]Department of Biomedical Engineering, Carnegie Mellon University, Pittsburgh, PA, USA. [7]These authors contributed equally: Adam Kohn, Christian K. Machens, Byron M. Yu. ✉e-mail: byronyu@cmu.edu

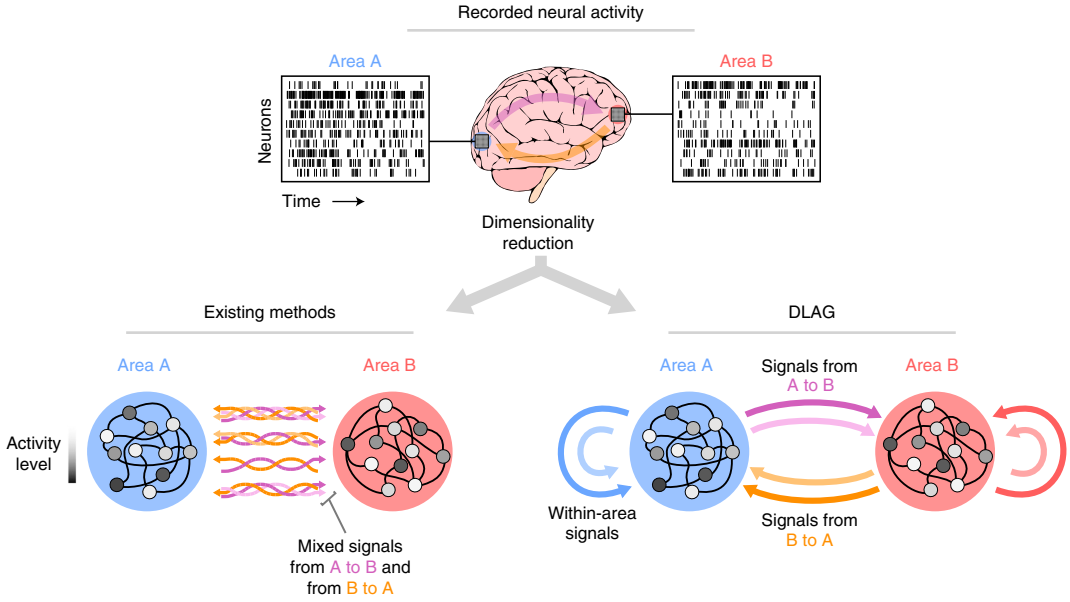

Fig. 1 | Disentangling the flow of signals between populations of neurons. Top: recorded neural activity provides only a tangled view of the bidirectional, concurrent interactions between brain areas (illustrated by the translucent magenta and orange arrows). Bottom left: existing dimensionality reduction methods identify correlated population activity across areas (each correlated population activity pattern is represented by a braid of multicoloured arrows; four different activity patterns are shown). Each activity pattern probably reflects a mixture of signals relayed in each direction. Within each activity pattern, individual arrows represent a directed interaction; colour depicts the direction of signal flow and shading (light versus dark) distinguishes distinct signals. Bottom right: DLAG identifies both within- and across-area population signals (indicated by colour and source/target of each arrow). Importantly, DLAG disentangles signals relayed in each direction. The colour of each arrow depicts the direction of signal flow associated with a population activity pattern, and shading distinguishes distinct signals.

Results

Model overview. Consider recording the activity of two populations of neurons (Fig. 2, left column), measured as, for example, the number of spikes counted within non-overlapping time bins. Here we will take these populations as belonging to two different brain areas, A and B. In principle, they can belong to any meaningful groups, such as cortical layers or cell types.

DLAG dissects the recorded population activity in each area on individual trials into a linear combination (weighted sum) of two types of latent variable (Fig. 2, centre column; equations (1) and (2) of Methods). The first type of latent variable, across-area variables, describes population activity that is correlated across areas (illustrated by the magenta box spanning both areas in Fig. 2). The second type of latent variable, within-area variables, describes population activity in one area that is not related to population activity in the other area (Fig. 2: blue, within A; red, within B). Whether or not the within-area variables are a subject of scientific study, they are critical to the correct estimation of across-area variables (Methods and Supplementary Discussion).

The temporal structure of within- and across-area variables are both described by relating each latent variable at different time points through Gaussian processes (GPs; equations (3)–(8) of Methods and Supplementary Fig. 1). Each GP is associated with its own characteristic timescale that controls the temporal smoothing of neural activity. Across-area variables are defined in pairs, where the elements of each pair correspond to the two areas and covary with each other according to a common GP (equation (6) of Methods). Importantly, the elements of each across-area pair are time delayed relative to each other (Fig. 2, D_1 between the first pair and D_2 between the second pair; equations (7) and (8) of Methods; Supplementary Fig. 1). The numbers of within- and across-area variables (that is, within- and across-area dimensionality) are estimated from the neural activity.

All DLAG model parameters, including the GP timescales and time delays, are estimated from the neural activity using an exact expectation-maximization algorithm (Supplementary Note). After the DLAG model parameters are estimated from the neural activity, the time courses of within- and across-area latent variables can be studied on a trial-to-trial basis. Conceptually, DLAG can be viewed as a time-series extension of probabilistic canonical correlation analysis (CCA)[42,43] or a multiarea extension of Gaussian process factor analysis (GPFA)[44,45] with the added ability to estimate time delays between two areas.

Intuitively, if a particular time course is reflected in the population activity of area A, and a similar time course, but after a time delay, is reflected in the population activity of area B, then an across-area variable pair can describe the apparent flow of that signal from A to B. Moreover, if concurrently a time course is first seen in area B, followed by area A, a second across-area variable pair can also describe the flow of that inter-areal signal. The key to disambiguating the first and second across-area variable pairs is that they involve different population activity patterns (that is, a 'loading' vector indicating how the activity of each neuron relates to the latent variable; equation (1) of Methods). In fact, DLAG can identify many across-area variable pairs, each with a delay of its own sign and magnitude, to capture multiple concurrent streams of signal flow between the two populations at different timescales.

The relationship between within- and across-area latent variables and observed population activity in each area can be represented geometrically with the concept of a population activity space (Fig. 2, right column). For each area, we can define a high-dimensional population activity space where each axis represents the activity of one neuron. Each point in the space represents the population activity at a particular time, and the points trace out a trajectory over time. DLAG's two types of latent variable each define the axes (dimensions) of a low-dimensional subspace within this population

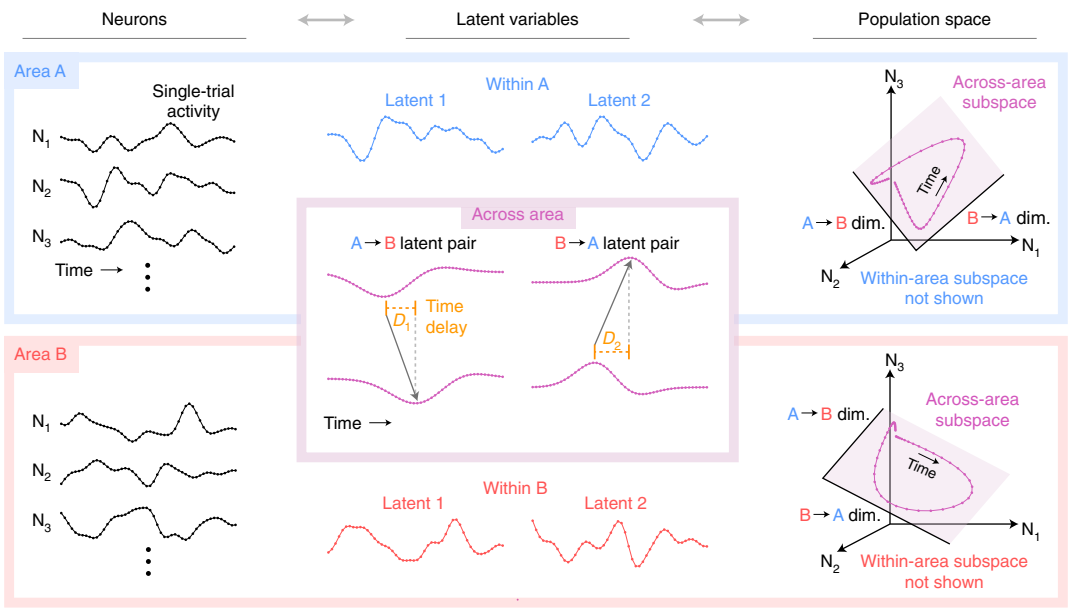

Fig. 2 | DLAG conceptual illustration. From left to right: neurons, latent variables and population activity space representations in two recorded brain areas analysed by DLAG. Left column: single-trial activity of neurons simultaneously recorded in each area. Only three neurons (N_1, N_2, N_3) are shown in each area for clarity. Centre column: within-area variables are shown in the colour corresponding to the area in which they belong. For clarity, only two within-area variables are shown in each area, but in principle there may be a greater number, as determined by DLAG from the recorded activity. Across-area variables are shown in magenta. The magenta box inset overlaps the blue and red boxes for areas A and B, respectively, to indicate that across-area variables are shared among neurons in both areas. The location of each across-area variable (that is, within the bounds of area A's box or area B's box) indicates which area's activity it reflects. Between area A and area B, across-area variables are vertically paired. The time courses of each pair are related after a time delay. The sign of this delay allows each pair to be associated with a directed interaction (A to B or B to A), which is indicated by grey arrows. For clarity, only two across-area variable pairs are shown. Right column: the activity of each neural population can be represented in a population activity space, where each axis represents the activity of a single neuron (N_1, N_2, N_3). Each point in population space represents the population activity at a particular time, and the points trace out a trajectory over time (magenta curve). DLAG identifies two linearly independent subspaces in each area: a within-area subspace (not shown, for clarity) and an across-area subspace (magenta-shaded plane). Each dimension ('dim.') of the across-area subspace is associated with a directed interaction.

activity space (in Fig. 2, we show only the across-area subspaces for visual clarity). Each dimension of these subspaces represents a population activity pattern.

Validation on realistic-scale synthetic data. Before applying DLAG to experimental data, we characterized its performance on synthetic datasets similar in scale to state-of-the-art neurophysiological recordings from multiple brain areas, and on additional synthetic datasets covering a wider range of experimental conditions. Informed by our recordings in macaque V1 and V2[26,36] (Dissecting bidirectional interactions between V1 and V2), we simulated independent datasets with representative numbers of neurons (area A, 80; area B, 20), trial counts (100), trial lengths (1,000 ms) and levels of noise, where noise is defined as the variance independent to each neuron (see Methods for additional details).

Across all datasets, within- and across-area latent time courses (Fig. 3a; see legend for quantification), across-area parameters (Fig. 3b, dimensionalities; Fig. 3c, delays; Fig. 3d, GP timescales) and within-area parameters (Fig. 3e, dimensionalities; Fig. 3f,g, GP timescales) were all consistently and accurately estimated. We highlight, in particular, DLAG's ability to estimate time delays between the two areas (Fig. 3c). Delay error was 1.3 ± 0.1 ms (mean and s.e.m. across all delays; maximum error 7.0 ms), despite observations occurring at 20 ms time steps. This accuracy emphasizes an important feature of the DLAG model that distinguishes it from other time-series modelling approaches (Discussion). Because latent time courses and time delays are continuous valued, DLAG can leverage the correlated activity of the neuronal populations to recover delays that are smaller than the sampling period (that is, spike count bin width, in the case of spiking activity).

The synthetic datasets presented here were generated with a variety of parameters representative of realistic data, but we also verified that DLAG performed well over a wider range of simulated conditions. Specifically, we systematically characterized DLAG's performance as a function of number of trials, number of neurons, latent dimensionality and noise level (Supplementary Fig. 2), as well as latent timescale (Supplementary Fig. 3). We also characterized the runtime of the DLAG fitting procedure as a function of number of trials, number of neurons, trial length and latent dimensionality (Supplementary Fig. 4).

Finally, we explored DLAG's robustness under several more challenging synthetic scenarios. DLAG's parameter and latent-variable estimates remained stable in instances where we induced imperfect estimates of dimensionality (Supplementary Figs. 5 and 6). DLAG also showed robustness to mild deviations from its assumptions of linearity and Gaussian observation noise (Supplementary Fig. 7; synthetic datasets were generated via a linear–nonlinear–Poisson model) and its assumption that neural activity follows a GP (Supplementary Fig. 8).

Dissecting bidirectional interactions between V1 and V2. We then used DLAG to study interactions between two areas in the early visual system: V1 and V2. V1 and V2 share strong reciprocal connections[46,47] and show correlated activity[22–24,26,36], but the bidirectional nature of their interactions is not yet well understood. We simultaneously recorded the activity of neuronal populations in the

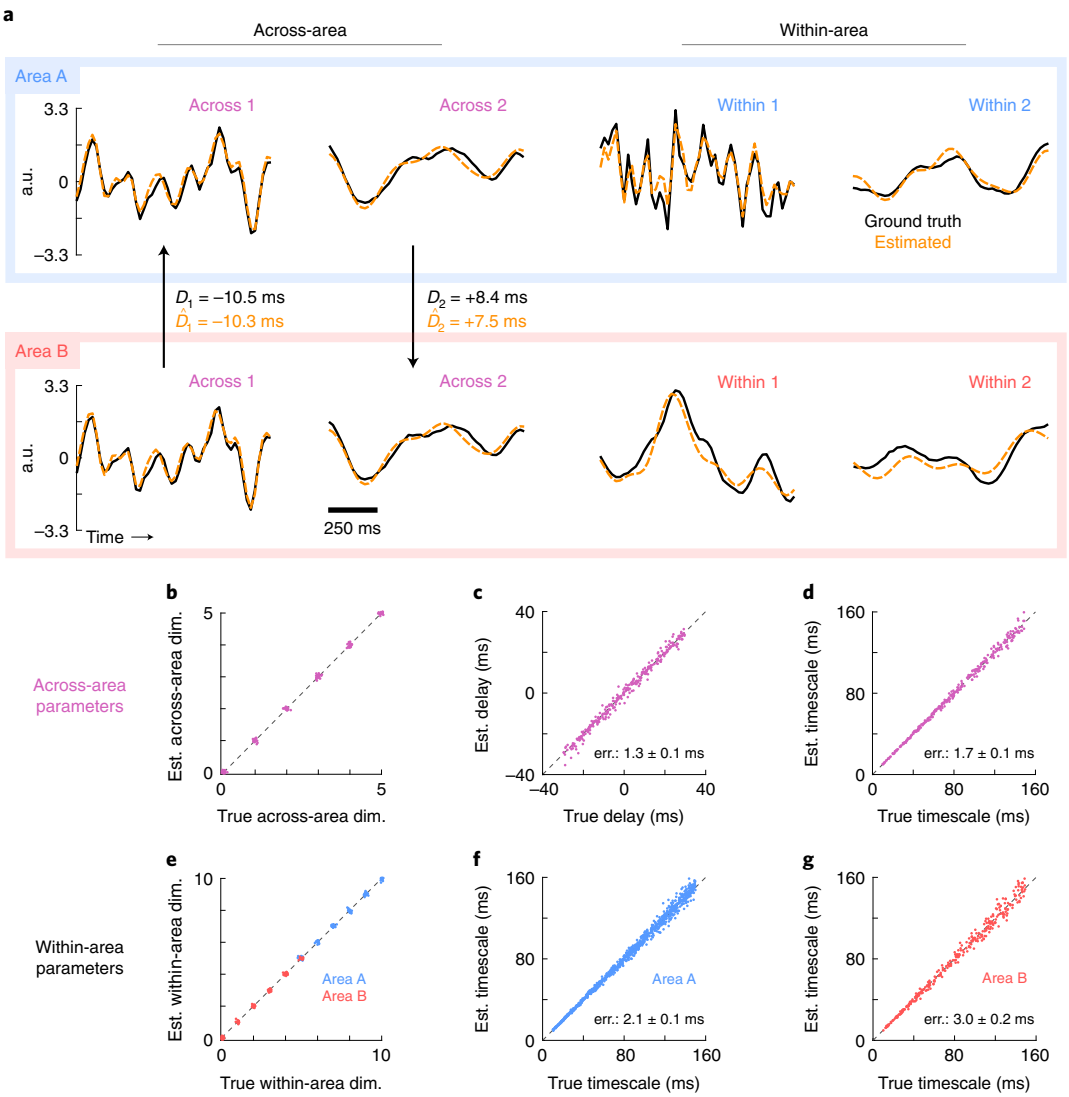

Fig. 3 | Estimates of within- and across-area time courses and their parameters in synthetic data. a, Single-trial latent-variable time-course estimates for a representative synthetic dataset. For visual clarity, two latent variables of each type (across-area and within-area) are shown. Orange dashed traces, DLAG estimates; black solid traces, ground truth. a.u., arbitrary units. Across all synthetic datasets for which across- or within-area dimensionality was non-zero (Methods; across, 100 datasets; within A, 120 datasets; within B, 100 datasets), the mean accuracy (R^2) of time-course estimation was as follows: area A, across—0.90; area B, across—0.91; area A, within—0.88; area B, within—0.82 (all s.e.m. values less than 0.01). Similarly, the mean accuracy of subspace (loading matrix) estimation was as follows: C_1^a—0.89; C_2^a—0.93; C_1^w—0.92; C_2^w—0.94 (where a value of 1 implies that the ground truth is fully captured by estimates; all s.e.m. 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 s.e.m. reported for 300 across-area variables. **d**, Across-area GP timescale estimates versus the ground truth. Displayed error indicates mean absolute error and s.e.m. reported for 300 across-area variables. **e**, Within-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. **f**, Within-area-A GP timescale estimates versus the ground truth. Displayed error indicates mean absolute error and s.e.m. reported for 900 within-area variables in area A. **g**, Within-area-B GP timescale estimates versus the ground truth. Displayed error indicates mean absolute error and s.e.m. reported for 300 within-area variables in area B.

superficial (output) layers of V1 (61–122 neurons; mean 86.3), and the middle (input) layers of V2 (15–32 neurons; mean 19.6) in three anaesthetized monkeys (Fig. 4a; data reported previously in refs. [26,36]). Recording locations were selected to maximize the probability that the recorded V1 and V2 populations interact by ensuring spatial receptive field alignment. We analysed neuronal responses measured during the 1.28 s presentation of drifting sinusoidal gratings of different orientations, and counted spikes in 20 ms time bins. The periodic nature of the drifting gratings (160 ms per cycle) is evident in peristimulus time histograms (PSTHs) for an example recording session and grating orientation (Fig. 4b). In total, we fitted DLAG models separately to 40 'datasets', corresponding to five recording sessions, each with eight different orientations. For comparison, on each dataset we also randomly split V1 into two equally sized subpopulations (termed V1a and V1b; Fig. 4c), and then applied DLAG to study V1a–V1b interactions in a manner identical to V1–V2.

We first used DLAG to study whether V1 and V2 interact selectively: whether in addition to fluctuations shared between V1 and V2 there are fluctuations that are not shared between the two areas. Selective inter-areal communication may be a hallmark of cortical computation that remains to be fully understood, particularly at the level of neuronal populations[5]. Indeed, significant across- and

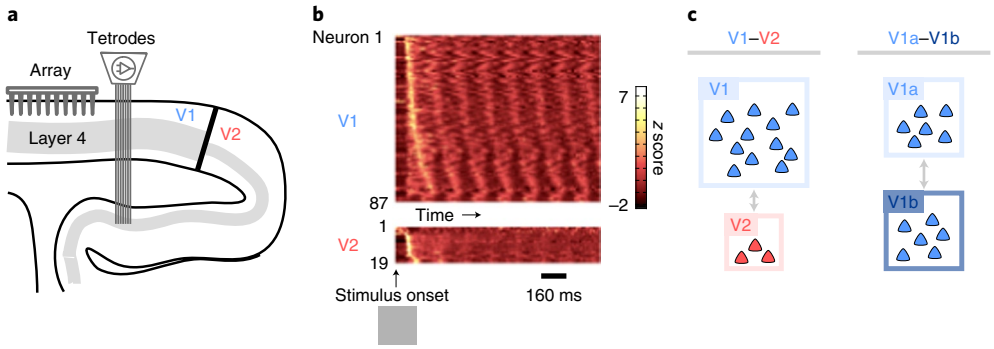

Fig. 4 | Simultaneous population recordings in V1 and V2. a, Schematic showing a sagittal section of occipital cortex and the recording set-up. V1 population activity was recorded using a 96-channel Utah array. V2 population activity was recorded using a set of movable electrodes and tetrodes. **b**, PSTHs during the stimulus presentation period, for an example session and stimulus condition. For visualization purposes, neuronal spike trains were first smoothed using a sliding Gaussian window of width 20 ms, and then *z* scored to produce normalized firing rates. Neurons are ordered from top to bottom (separately for V1 and V2) according to the time at which their peak firing rate occurs. **c**, Inter- and intra-areal comparisons. Left: we applied DLAG to spike counts in V1 and V2. Right: for comparison, we applied DLAG to two equally sized V1 subpopulations (V1a and V1b), randomly selected from the V1 population. Each triangle represents a neuron. Box sizes illustrate typical relative population sizes.

within-area latent variables (that is, latent variables that were selected via cross-validation) were identified consistently across datasets (Fig. 5a: single-trial latent time courses from a representative dataset; Fig. 6a, top, dimensionalities across all datasets; median dimensionality across areas, 3; within V1, 14; within V2, 2).

We further sought to characterize the strength (in addition to the dimensionality) of across- versus within-area activity in each area. We therefore considered the latent variables in V1 and in V2 separately, and computed the fraction of shared variance that each latent variable explained in its corresponding area (in Fig. 5, the amplitude of each latent time course is scaled by this value). Across-area variables explained only a portion of the shared variance in V1 and in V2 (Fig. 6b, top; median across-area strengths, 34% in V1; 76% in V2). Interestingly, across-area activity explained more of the shared variance in V2 than in V1 (Fig. 6b, top, points above the diagonal). This observation could not be fully attributed to differences in recorded population size or in the total dimensionality of each area (Supplementary Fig. 9). This difference in across-area strength might be a consequence of the cortical layers from which we recorded: much of the activity in the middle layers of V2 is probably driven by V1. The superficial layers of V1, on the other hand, receive input from other sources that do not also project to the middle layers of V2.

Collectively, these observations (Fig. 6a,b, top) are consistent with the presence of a communication subspace between V1 and V2[36], through which only a subset of population activity patterns is shared between the two areas. Our results further suggest that not only does there exist activity in V1 that is not shared with V2 (as reported in ref. [36]), but there also exists activity in V2 that is not shared with V1. By contrast, V1a and V1b do not interact selectively. V1a–V1b 'across-population' activity was of higher dimensionality than 'within-population' activity and V1–V2 across-area activity (Fig. 6a, bottom; median dimensionality across populations, 11; within V1a, 2; within V1b, 1), and accounted for nearly all of the shared variance in V1a and in V1b (Fig. 6b, bottom; median across-population strengths, 96% in V1a, 98% in V1b; note also the small amplitudes of the within-population latent time courses in Fig. 5b).

DLAG's latent variables enabled further qualitative characterization of the moment-to-moment nature of within- and across-area activity on individual trials. For instance, stereotyped periodic signals, whose periods matched the period of the drifting grating presented, appeared strongly within V1 (Fig. 5a, top, 'Across 3', 'Within 1' and 'Within 2') and only weakly in V2 (Fig. 5a, bottom, Across 3). The prominence of this stimulus-related periodic structure in V1 relative to V2 is consistent with the stimulus response properties of neurons in each area[48], evident in the neuronal PSTHs (Fig. 4b). Care should be taken, however, when interpreting these latent variables as across-area interactions (Discussion). By contrast, periodic signals were not evident in V1a or V1b within-population variables, but were evident in the activity shared between V1a and V1b (Fig. 5b, 'Across 1' and 'Across 2'). Other latent variables, particularly within V2, exhibited additional trial-to-trial variability whose connection to the presented stimulus is less apparent (for example, Fig. 5a, bottom, Within 1 and Within 2). Latent-variable time courses, as well as estimated across- and within-area subspaces, were also sensitive to the orientation of presented stimuli in a manner consistent with the tuning of neurons in V1 and V2 (Supplementary Fig. 10). DLAG also detected differences in the type of stimulus presented (oriented gratings versus naturalistic textures) when we used it to study V1–V4 interactions in an awake animal (Supplementary Fig. 11).

We next used DLAG to study the bidirectional nature of interactions between V1 and V2. Note that this task may not be straightforward with a static dimensionality reduction method such as CCA (Supplementary Figs. 12–14). Each of DLAG's across-area latent variables is associated with a time delay that indicates a feedforward (positive delay: V1 to V2) or feedback (negative delay: V2 to V1) interaction. For example, the first representative V1–V2 across-area variable (Fig. 5a, Across 1) was associated with a −23 ms delay, implying a feedback interaction. In contrast, the visually similar V1a–V1b across-population variable (Fig. 5b, Across 3) was associated with a 0 ms delay. A V1a–V1b delay at or near zero is expected, given that the V1a and V1b populations belong to the same area, and probably receive common inputs with similar latencies (in contrast to the populations in distinct areas V1 and V2).

We developed a statistical procedure to test whether such delays significantly deviate from zero. In brief, we assessed whether setting the delay to 0 ms resulted in a significant reduction in model performance; if so, the delay was deemed significant (that is, 'non-zero'; Methods). Indeed, the directionality of this latent variable (Across 3 for V1a–V1b) was identified as statistically 'ambiguous' (that is, not significantly different from zero, indicated by the bidirectional grey arrow in Fig. 5b). In separate analyses, we also verified that V1–V2 interactions are better described by DLAG models with time delays than without time delays (Supplementary Fig. 15).

Delays across all datasets reflected bidirectional interactions between V1 and V2 (Fig. 6c, top). Notably, the delays between V1 and V2 exhibited a striking asymmetry. The interactions across

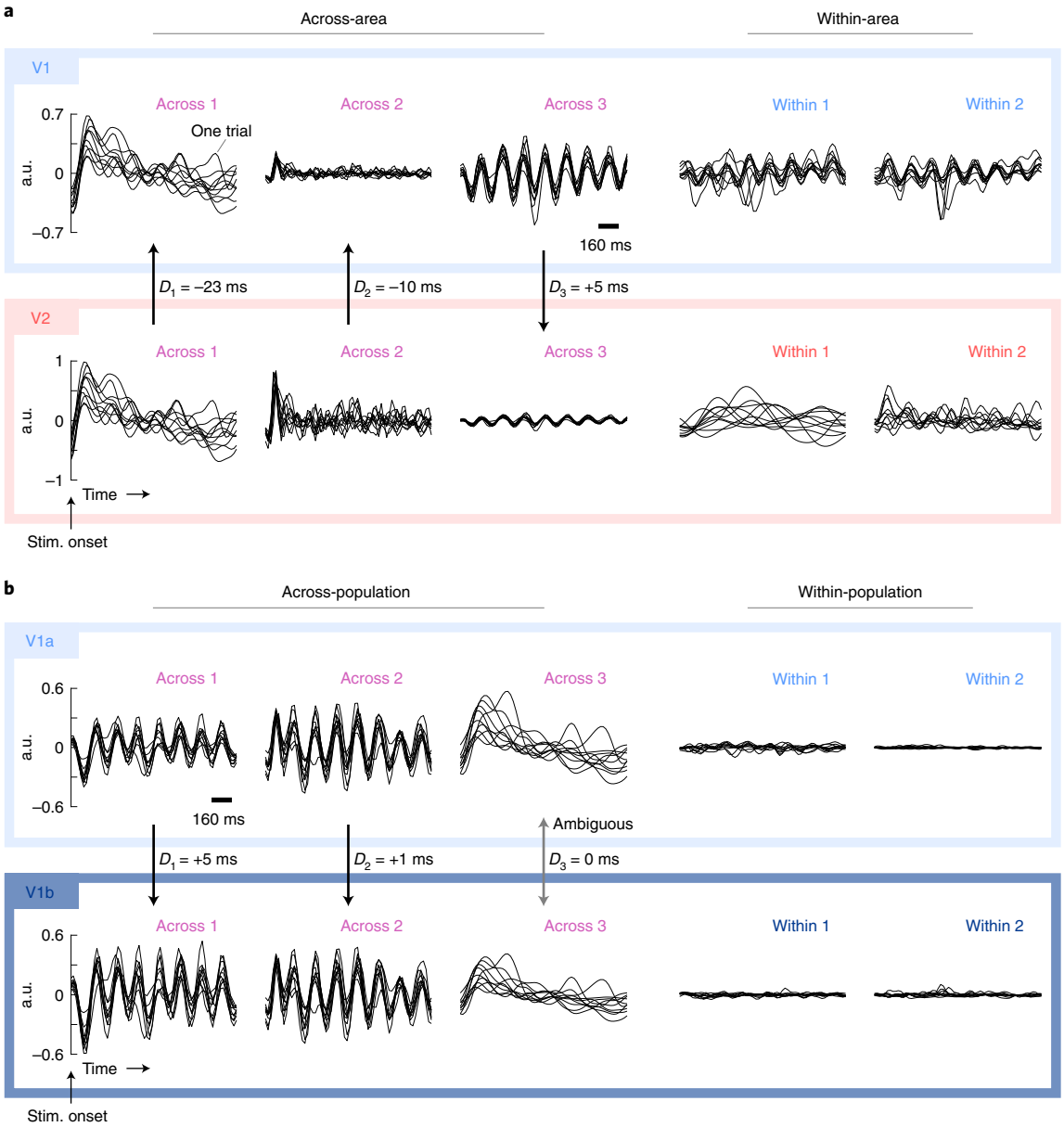

Fig. 5 | Representative DLAG time courses for inter- and intra-areal analyses (same dataset as shown in Fig. 4b). a, V1–V2 time courses. Each panel corresponds to the single-trial time courses of a latent variable. All time courses are aligned to stimulus onset. Each black trace corresponds to one trial; for clarity, only 10 of 400 are shown. Note that the polarity of traces is arbitrary, as long as it is consistent with the polarity of C_1^a or C_1^w. Across-area variables are paired vertically; vertical arrows point in the direction of the identified signal flow, as determined by the sign of the delay next to each arrow. All delays for the displayed dataset were deemed significantly different from zero (Methods). For visualization purposes, latent variables have been scaled and ordered by the fraction of shared variance they explain (across- and within-area variables are sorted separately; across-area variables are sorted according to shared variance explained in V2). All across-area variables and within-V2 variables uncovered by DLAG are shown here. The top 2 of 14 within-V1 variables are displayed, which explain 46% of V1's within-area shared variance. **b**, V1a–V1b time courses. Conventions are the same as in **a**. Here, the delay for the third across-population variable (Across 3) was deemed to have an ambiguous sign, indicated by the bidirectional grey arrow. All other delays for the displayed dataset were deemed significantly different from zero, indicated by the unidirectional black arrows. Three of 10 across-population variables uncovered by DLAG, which explain 23% and 17% of V1a's and V1b's total shared variance, respectively, are shown here. All uncovered within-V1a variables are shown, and two of five within-V1b variables are shown. Within-population variables (including those not shown here) explained 5% and 7% of V1a's and V1b's total shared variance, respectively.

these areas were predominantly directed from V2 to V1 (Fig. 6c, top; median over non-zero delays, −8 ms; median over all delays, −5 ms). Among the across-area latent variables with statistically significant delays, 76% were associated with a negative delay. This asymmetry remained even when we subsampled the V1 population to match V2 in size, and reapplied DLAG (Supplementary Fig. 9). Similarly to the strength of across-area activity observed in V1 and in V2 (Fig. 6b, top), the magnitudes of the delays might also reflect the cortical layers from which we recorded. The positive delays tended to be short (Fig. 6c, top; median across significant positive delays, +7 ms), consistent with the fact that the superficial layers of V1 directly project to the middle layers of V2[24,26]. The negative delays tended to be longer (Fig. 6c, top; median across significant negative delays, −11 ms), consistent with a multisynaptic

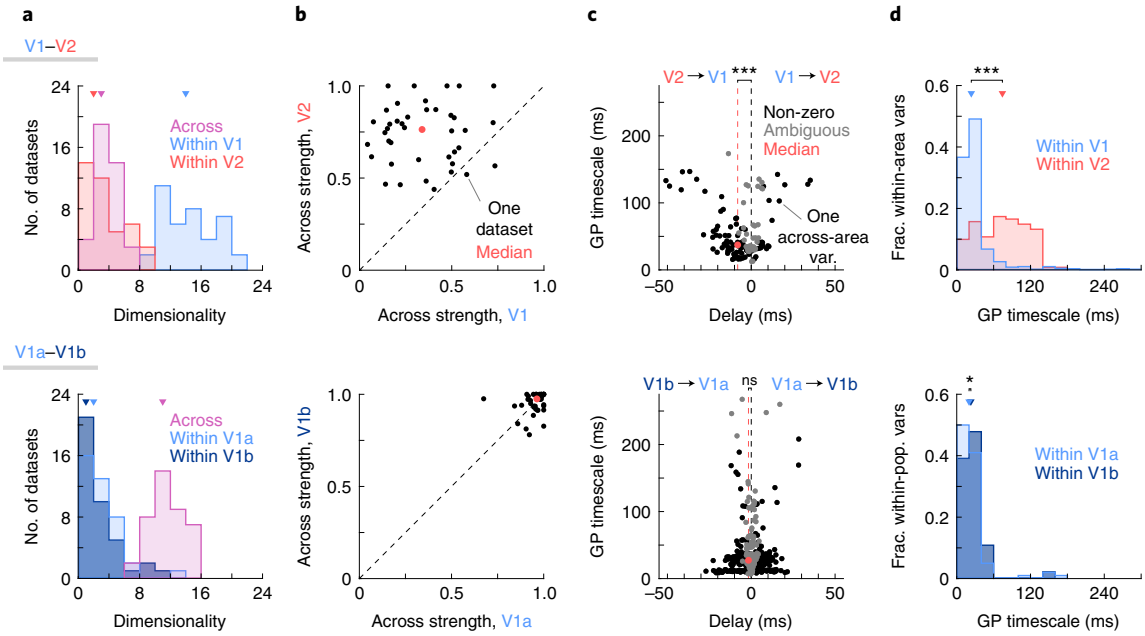

Fig. 6 | Uncovering properties of V1–V2 interactions with DLAG. a, Within- and across-area dimensionalities (determined via cross-validation). Top: V1–V2 results. Distribution of within-V1, within-V2 and across-area dimensionalities across 40 datasets. Triangles indicate the median of each distribution. Bottom: V1a–V1b results; same format. **b**, Fraction of shared variance of each area explained by across-area latent variables. Top: V1–V2 results. Across-area strength is significantly greater in V2 than in V1 (one-sided paired sign test; $P = 7.5 \times 10^{-10}$). Bottom: V1a–V1b results; same format. Across-population strength is not significantly greater in one population or the other (two-sided paired sign test; $P = 0.868$). **c**, GP timescale versus time delay for across-area latent variables. Top: V1–V2 results. Each point represents one across-area latent variable. Black points, across-area latent variables for which the delays were deemed significantly non-zero (Methods; 95 of 135 across-area variables across all 40 datasets). Grey points: across-area latent variables for which delays were deemed ambiguous (not significantly positive or negative; 40 of 135 across-area variables across all 40 datasets). ***Delays are significantly less than zero, representing feedback interactions from V2 to V1 (one-sided one-sample sign test on non-zero delays, $P = 2.4 \times 10^{-7}$). Bottom: V1a–V1b results; same format. Out of 437 across-population latent variables uncovered across all 40 datasets, 316 delays were deemed significantly non-zero, while 121 delays were deemed ambiguous. ns, delays are not significantly negative (one-sided one-sample sign test on non-zero delays, $P = 0.08$). **d**, GP timescales for within-area latent variables. Top: V1–V2 results. Normalized distribution of within-V1 and within-V2 GP timescales across all 40 datasets (total within-V1 latent variables, 562; total within-V2 latent variables, 121). Triangles indicate the median of each distribution. ***Within-V2 GP timescales are significantly longer than within-V1 GP timescales (one-sided Wilcoxon rank sum test, $P = 1.6 \times 10^{-31}$). Bottom: V1a–V1b results; same format (total within-V1a latent variables, 100; total within-V1b latent variables, 92). *Within-V1b GP timescales are significantly longer than within-V1a GP timescales (one-sided Wilcoxon rank sum test, $P = 0.039$), even though the magnitude of the difference is small (as expected for randomly assigned subpopulations).

path from the middle layers of V2 back to the superficial layers of V1. We also found that the strongest across-area interactions in V1 were nominally feedforward (V1 to V2), while the strongest across-area interactions in V2 were nominally feedback (V2 to V1) (Supplementary Fig. 16).

By contrast, V1a–V1b interactions were symmetric (Fig. 6c, bottom; median over non-zero delays, −2 ms; median over all delays, 0 ms; neither median significantly different from zero; 54% of non-zero delays were negative; see also Supplementary Fig. 16). This centring of the delay distribution around zero is expected, given that the neurons in V1a and V1b were randomly chosen and belong to the same area. Still, the magnitudes of V1a–V1b delays were not universally zero. These non-zero delays probably reflect aggregate differences in the stimulus response properties of the randomly chosen V1a and V1b subpopulations. For example, inspection of PSTHs (Fig. 4b) suggests that the phase of trial-averaged periodic structure can vary by tens of milliseconds between individual V1 neurons.

Finally, we examined the timescales of neural activity identified by DLAG within V1 and V2. Within-V2 GP timescales were longer than within-V1 GP timescales (Fig. 6d, top; median within V1, 24 ms; within V2, 74 ms). Within-V1a and within-V1b GP timescales, on the other hand, were nearly the same (Fig. 6d, bottom; median within V1a, 20 ms; within V1b, 23 ms). These observations are consistent with previous evidence that timescales increase for areas higher up the cortical hierarchy[49–51].

Discussion

By leveraging the correlated activity across two neuronal populations, DLAG can disentangle concurrent signals relayed in each direction and characterize how those signals evolve over time within and across trials. Although we applied DLAG to the spiking activity of populations of neurons in distinct brain areas, DLAG is applicable to any high-dimensional time-series data, including other neural recording modalities (for example, calcium imaging, subject to the temporal resolution inherent to the recording technology). It can also be used to study the interaction of two populations of neurons in different cortical layers or of different cell types. DLAG can even be used to study the relationship between a neuronal population and a dynamic stimulus or behavioural variables.

Recently, feedforward and feedback signalling was studied in the same V1–V2 recordings as analysed here[41]. CCA was used in a sliding window scheme to identify trial epochs dominated by either feedforward or feedback signalling. V1–V2 (and V1–V4) interactions were found to involve distinct population activity patterns during feedforward- versus feedback-dominated trial epochs. This statistical approach, however, could not be used to study the concurrent nature of feedforward and feedback signalling (see Supplementary

Figs. 13 and 14 for further discussion). In this Article, we have provided a complementary view of V1–V2 interactions, using DLAG to identify concurrent, distinct feedforward and feedback activity patterns that characterize the stimulus presentation period as a whole. Future work could characterize how the activity patterns uncovered by DLAG and their associated time delays might change during the course of a trial (see more below).

DLAG identified V1–V2 interactions in both directions, whose strengths and associated time delays appear to reflect the cortical layers from which we recorded. DLAG might have been expected to identify at least as many feedforward (V1 to V2) interactions as feedback (V2 to V1). Generally, feedback intercortical connections equal feedforward connections in number; moreover, specific to our recording arrangement, feedback connections do not originate in the input layers of V2[46,47]. Surprisingly, DLAG revealed a marked asymmetry, such that a majority of across-area latent variables were associated with a feedback interaction. This apparent disparity presents an opportunity for future study.

DLAG has commonalities with several other methods. For instance, static dimensionality reduction methods such as CCA, sparse structured CCA and their probabilistic variants[52] identify across- and/or within-area latent variables, but do not characterize inter-areal interactions over time or the directionality of signal flow (but see ref. [41], discussed above). Multivariate time-series methods such as Granger causal modelling[53–55], generalized linear models[37,56,57] or recurrent neural networks[58] characterize the directionality of signal flow, but not in a low-dimensional manner. Time-series methods that provide a low-dimensional description of across-area activity do not provide a low-dimensional description of within-area activity, should low-dimensional within-area activity be of scientific interest[59,60]; or, they do not characterize time delays between areas[61]. In contrast with all of these methods, DLAG jointly reduces dimensionality and characterizes the directionality of signal flow by estimating across- and within-area latent variables with time delays and timescales.

DLAG offers unique advantages when characterizing the temporal structure of activity within and across areas. Applied to V1 and V2, DLAG uncovered latent variables with diverse temporal profiles and timescales. The ability to capture diverse dynamical motifs stems from DLAG's definition via GPs[44]: beyond temporal smoothness, DLAG makes no additional assumptions about the form of dynamics within or across areas. In contrast, multiarea methods proposed by refs. [62] and [63], for instance, describe interactions between areas according to a parametric dynamical model. GPs provide DLAG with another advantage: the ability to discover wide-ranging delays with high precision[45]. Existing multiarea methods (nearly all of which, above, are defined in discrete time) are limited to delays restricted to be integer multiples of the sampling period or spike count bin width of neural activity.

With the conceptual and statistical advantages described above, DLAG is a powerful tool for exploratory data analysis. For example, after performing a new experiment, one can use DLAG to generate data-driven hypotheses about plausible dynamical motifs within and across areas. Then, one can test these hypotheses using a dynamical system-based approach—for example, data-constrained recurrent networks[58,62,63].

The population activity patterns represented by DLAG's across-area variables might be interpreted as distinct 'channels' through which two areas communicate[41]. As with any statistical method, however, interpretation of the features extracted by DLAG is subject to ambiguities, particularly when not all relevant brain areas and neurons are recorded[32,64]. An across-area latent variable, for instance, could reflect an interaction between areas A and B that is either direct or indirect, mediated by a third (unobserved) area C. Similarly, a within-area latent variable could reflect activity internal to one area, or it could reflect inputs sent from unrecorded neurons to one area but not the other.

The sign and magnitude of DLAG's time delays can, however, narrow the set of hypotheses consistent with the data. We might reasonably suspect, for example, that short positive (V1 to V2) delays identified by DLAG reflect direct interactions from the output layers of V1 to the input layers of V2 (the layers from which we recorded)[24,26]. Larger negative (V2 to V1) delays might instead indicate indirect interactions, given that the path from the input layers of V2 to the output layers of V1 involves multiple synapses. Some across-area latent variables were associated with delays statistically indistinguishable from zero (that is, ambiguous), and could indicate either tight recurrent interactions or common input from an unobserved source. Future experimental interventions could further disambiguate these cases.

A phenomenon widely recognized by cross-correlation studies[21–26] is the presence of correlations across areas due simply to common stimulus drive, rather than an inter-areal interaction. For DLAG, these stimulus-driven effects can appear as an across-area variable. The stereotyped periodic signals evident in V1–V2 across-area latent variables (Fig. 5a; Across 3) are a probable example. If desired, one could control for these effects with straightforward preprocessing steps, such as the subtraction of PSTHs from single-trial responses, thereby emphasizing trial-to-trial fluctuations correlated across areas[36].

Assumptions explicit in the DLAG model definition warrant additional care when interpreting estimated delays. First, DLAG treats time delays as constant parameters. However, the direction of interaction associated with a dimension of population activity might not be constant across different trial epochs or different experimental (for example, stimulus) conditions. Thus, we interpret a delay as a summary of this direction of interaction throughout the course of an experiment. Similarly, neurons within the same area can respond to a common input with different latencies (evident in, for example, Fig. 4b). An estimated delay hence also represents a summary across neurons[45]. Second, DLAG assumes that each dimension of population activity is associated with one delay, or direction. If a set of interactions were to occur concurrently in both directions but evolve along the same dimension, then teasing apart directionality might be difficult—albeit for any statistical method, not just DLAG. Third, DLAG assumes that signals are linearly transformed across areas. DLAG therefore does not take into account nonlinear transformations of signals. We believe that there are many experimental scenarios for which the assumption of a linear transformation or direct signal transmission is appropriate (for example, ref. [14]; Supplementary Fig. 7). Nonetheless, in practice, this assumption should be evaluated on a case-by-case basis.

Solutions to these interpretational challenges might be well within reach, if not already available through DLAG's existing machinery. For example, one could fit DLAG to subsets of trials, to subsets of neurons or to separate trial epochs to understand how DLAG's estimates depend on these elements of the neural recordings. We have already employed some of these strategies here (Fig. 6 and Supplementary Figs. 9 and 11), and could continue to build upon this foundation.

Methods

Mathematical notation. To disambiguate each variable or parameter in the DLAG model, we need to keep track of up to four labels that indicate their associated (1) subpopulation (for example, brain area), (2) neuron or latent variable index, (3) time point or (4) designation as within- or across-area. We indicate the first three labels via subscripts, where subpopulations (areas) are indexed by $i = 1, 2$, neurons or latent variables are indexed by j (we will indicate the upper bound as appropriate) and time is indexed by $t = 1, \ldots, T$. For example, we define the observed activity of neuron j (out of q_i) in area i at time t as $y_{i,j,t} \in \mathbb{R}$. To indicate a collection of all variables along a particular index, we replace that index with a colon. Hence we represent the simultaneous activity of the population of q_i neurons observed in area i at time t as the vector $\mathbf{y}_{i,:,t} \in \mathbb{R}^{q_i}$. For concision, where a particular index is either not applicable or not immediately relevant, we omit it. The identities of the remaining indices should be clear from context. For example,

throughout this work we consider only the activity of a full population, and not of single neurons, so we rewrite $\mathbf{y}_{i,:,t}$ as $\mathbf{y}_{i,t}$. Finally, we indicate a latent variable's or parameter's designation as within- or across-area via a superscript, where 'w' indicates within-area, and 'a' indicates across-area. For example, we define across-area latent variable j (out of p^{a}) in area i at time t as $x^{\mathrm{a}}_{i,j,t} \in \mathbb{R}$, and the collection of all p^{a} latent variables as the vector $\mathbf{x}^{\mathrm{a}}_{i,:,t} \in \mathbb{R}^{p^{\mathrm{a}}}$. We similarly define within-area latent variable j (out of p^{w}_i) in area i at time t as $x^{\mathrm{w}}_{i,j,t} \in \mathbb{R}$, and the collection of all p^{w}_i latent variables as the vector $\mathbf{x}^{\mathrm{w}}_{i,:,t} \in \mathbb{R}^{p^{\mathrm{w}}_i}$.

It is conceptually helpful to understand the above notation for observed ($\mathbf{y}$) and latent ($\mathbf{x}$) variables as taking cross-sections of matrices. For example, observed activity in area i can be grouped into the matrix $Y_i = [\mathbf{y}_{i,1} \cdots \mathbf{y}_{i,T}] \in \mathbb{R}^{q_i \times T}$. Then, each $\mathbf{y}_{i,t}$ is a column of Y_i. Similarly, across-area latent variables in area i can be grouped into the matrix $X^{\mathrm{a}}_i = [\mathbf{x}^{\mathrm{a}}_{i,:,1} \cdots \mathbf{x}^{\mathrm{a}}_{i,:,T}] \in \mathbb{R}^{p^{\mathrm{a}} \times T}$. Each $\mathbf{x}^{\mathrm{a}}_{i,:,t}$ is a column of X^{a}_i. Similarly, we represent a row of X^{a}_i (that is, the values of a single latent variable j at all time points) as $\mathbf{x}^{\mathrm{a}}_{i,j,:} \in \mathbb{R}^T$. Within-area latent variables can be understood analogously from the matrix $X^{\mathrm{w}}_i = [\mathbf{x}^{\mathrm{w}}_{i,:,1} \cdots \mathbf{x}^{\mathrm{w}}_{i,:,T}] \in \mathbb{R}^{p^{\mathrm{w}}_i \times T}$. Finally, we note that there is a separate set of observed and latent variables (Y_i, X^{a}_i, X^{w}_i) for each trial, while there is a single set of DLAG model parameters shared across trials. For concision, we index trial number only as needed, and omit the trial index otherwise.

We will explicitly define all other variables and parameters as they appear, but for reference we list common variables and parameters below.

Observed neural activity.

- q_i — number of neurons observed in area i
- Y_i — $q_i \times T$ matrix of observed activity in area i
- $\mathbf{y}_{i,t}$ — $q_i \times 1$ vector of observed activity in area i at time t; the t^{th} column of Y_i

Latent variables.

- p^{a} — number of across-area variables (same for both areas)
- X^{a}_i — $p^{\mathrm{a}} \times T$ matrix of across-area variables in area i
- $\mathbf{x}^{\mathrm{a}}_{i,:,t}$ — $p^{\mathrm{a}} \times 1$ vector of across-area variables in area i at time t; the t^{th} column of X^{a}_i
- $\mathbf{x}^{\mathrm{a}}_{i,j,:}$ — $T \times 1$ vector of values of across-area variable j in area i over time; the j^{th} row of X^{a}_i
- p^{w}_i — number of within-area variables in area i
- X^{w}_i — $p^{\mathrm{w}}_i \times T$ matrix of within-area variables in area i
- $\mathbf{x}^{\mathrm{w}}_{i,:,t}$ — $p^{\mathrm{w}}_i \times 1$ vector of within-area variables in area i at time t; the t^{th} column of X^{w}_i
- $\mathbf{x}^{\mathrm{w}}_{i,j,:}$ — $T \times 1$ vector of values of within-area variable j in area i over time; the j^{th} row of X^{w}_i

Model parameters.

- C^{a}_i — $q_i \times p^{\mathrm{a}}$ across-area loading matrix for area i
- C^{w}_i — $q_i \times p^{\mathrm{w}}_i$ within-area loading matrix for area i
- $\mathbf{d}_i$ — $q_i \times 1$ mean parameter for area i
- R_i — $q_i \times q_i$ observation noise covariance matrix for area i
- $D_{i,j}$ — time delay parameter between area i and across-area variable j
- D_j — relative time delay associated with across-area variable j; $D_j = D_{2,j} - D_{1,j}$
- τ^{a}_j — GP timescale for across-area variable j
- σ^{a}_j — GP noise parameter for across-area variable j
- $\tau^{\mathrm{w}}_{i,j}$ — GP timescale for within-area variable j in area i
- $\sigma^{\mathrm{w}}_{i,j}$ — GP noise parameter for within-area variable j in area i

GP covariances.

- $K^{\mathrm{a}}_{i_1,i_2,j}$ — $T \times T$ covariance matrix for across-area variable j, between areas i_1 and i_2
- $k^{\mathrm{a}}_{i_1,i_2,j}$ — covariance function for across-area variable j, between areas i_1 and i_2
- $K^{\mathrm{w}}_{i,j}$ — $T \times T$ covariance matrix for within-area variable j in area i
- $k^{\mathrm{w}}_{i,j}$ — covariance function for within-area variable j in area i

DLAG observation model. For area i at time t, we define a linear–Gaussian relationship between observed activity, $\mathbf{y}_{i,t}$, and latent variables, $\mathbf{x}^{\mathrm{a}}_{i,:,t}$ and $\mathbf{x}^{\mathrm{w}}_{i,:,t}$ (ref. [43]):

$$\mathbf{y}_{i,t} = C^{\mathrm{a}}_i \mathbf{x}^{\mathrm{a}}_{i,:,t} + C^{\mathrm{w}}_i \mathbf{x}^{\mathrm{w}}_{i,:,t} + \mathbf{d}_i + \boldsymbol{\varepsilon}_i \tag{1}$$

$$\boldsymbol{\varepsilon}_i \sim \mathcal{N}(0, R_i) \tag{2}$$

where $C^{\mathrm{a}}_i \in \mathbb{R}^{q_i \times p^{\mathrm{a}}}$, $C^{\mathrm{w}}_i \in \mathbb{R}^{q_i \times p^{\mathrm{w}}_i}$, $\mathbf{d}_i \in \mathbb{R}^{q_i}$ and $R_i \in \mathbb{S}^{q_i \times q_i}$ ($\mathbb{S}^{q_i \times q_i}$ is the set of $q_i \times q_i$ symmetric matrices) are model parameters to be estimated from data. The relationship between observed and latent variables is illustrated graphically in Supplementary Fig. 1a. The loading matrices C^{a}_i and C^{w}_i linearly combine latent variables and map them to observed neural activity. The parameter $\mathbf{d}_i$ can be thought of as the mean firing rate of each neuron. $\boldsymbol{\varepsilon}_i$ is a zero-mean Gaussian random variable, where we constrain the covariance matrix R_i to be diagonal, as in factor analysis (FA) and GPFA[44], to capture variance that is independent to each neuron. This constraint encourages the latent variables to explain as much of the shared variance among neurons as possible.

As we will describe, at time point t, across-area variables $\mathbf{x}^{\mathrm{a}}_{1,:,t}$ and $\mathbf{x}^{\mathrm{a}}_{2,:,t}$ in area 1 and area 2, respectively, are coupled with each other, and thus each area has the same number of across-area variables, p^{a}. Within-area variables are not coupled across areas, on the other hand, and thus each area i may have a different number of within-area variables, p^{w}_i. Because we seek a low-dimensional description of neural activity in each area, the combined number of across- and within-area variables is less than the number of neurons, that is, $p^{\mathrm{a}} + p^{\mathrm{w}}_i < q_i$, where p^{a} and p^{w}_i are determined by the data (see below).

The parameters C^{w}_i and C^{a}_i have an intuitive geometric interpretation (Fig. 2, right column). Each element of $\mathbf{y}_{i,t}$ can be represented as an axis in a high-dimensional population activity space. Then the columns of C^{a}_i, the across-area loading matrix for area i, define a subspace in this population activity space, where each dimension corresponds to a distinct across-area latent variable. This across-area subspace represents patterns of population activity that is correlated across areas. Analogously, the columns of C^{w}_i define a within-area subspace, which represents patterns of population activity that is shared only among neurons within area i. Additionally, as we will discuss below, since the j^{th} pair of across-area variables ($\mathbf{x}^{\mathrm{a}}_{1,j,:}$, $\mathbf{x}^{\mathrm{a}}_{2,j,:}$) is associated with a direction of population signal flow (Fig. 2, centre column), so too are the corresponding columns in C^{a}_1 and C^{a}_2. The across-area subspace can thus be partitioned further on the basis of the nominal directionality of activity patterns (area 1 to area 2, or area 2 to area 1). Finally, note that the columns of C^{a}_i are linearly independent but not, in general, orthogonal. Likewise, the columns of C^{w}_i are linearly independent but not, in general, orthogonal. The across- and within-area subspaces in area i (spanned by the columns of C^{a}_i and by the columns of C^{w}_i, respectively) are also linearly independent but not, in general, orthogonal. The ordering of the columns of each loading matrix, and of the corresponding latent variables, is arbitrary.

DLAG state model. We seek to extract smooth, single-trial latent time courses, where the degree of smoothing is determined by the neural activity (as described below). The time course of each within-area and across-area latent variable is described by a GP[65].

Within-area latent variables. For each within-area variable $j = 1, \ldots, p^{\mathrm{w}}_i$ in brain area i, we define a separate GP as follows[44]:

$$\mathbf{x}^{\mathrm{w}}_{i,j,:} \sim \mathcal{N}(0, K^{\mathrm{w}}_{i,j}) \tag{3}$$

where $K^{\mathrm{w}}_{i,j} \in \mathbb{S}^{T \times T}$ is the covariance matrix for within-area variable j of area i. DLAG is compatible with any valid form of GP covariance, but for the present work we choose the commonly used squared exponential function. Then, element (t_1, t_2) of $K^{\mathrm{w}}_{i,j}$, the covariance between samples of the within-area variable at times t_1 and t_2, can be computed according to

$$k^{\mathrm{w}}_{i,j}(t_1, t_2) = \left(1 - (\sigma^{\mathrm{w}}_{i,j})^2\right) \exp\left(-\frac{(\Delta t)^2}{2(\tau^{\mathrm{w}}_{i,j})^2}\right) + (\sigma^{\mathrm{w}}_{i,j})^2 \cdot \delta_{\Delta t} \tag{4}$$

$$\Delta t = t_2 - t_1 \tag{5}$$

where the characteristic timescale, $\tau^{\mathrm{w}}_{i,j} \in \mathbb{R}_{>0}$, and GP noise variance, $(\sigma^{\mathrm{w}}_{i,j})^2 \in (0, 1)$, are model parameters. $\delta_{\Delta t}$ is the Kronecker delta, which is 1 for $\Delta t = 0$ (equivalently, $t_1 = t_2$) and 0 otherwise.

Notice that $k^{\mathrm{w}}_{i,j}$ is stationary: the squared exponential function depends only on the time difference $(t_2 - t_1)$ (Supplementary Fig. 1b). This stationarity gives the covariance matrix $K^{\mathrm{w}}_{i,j}$ a characteristic banded structure (Supplementary Fig. 1c). The characteristic timescale, $\tau^{\mathrm{w}}_{i,j}$, dictates the width of $k^{\mathrm{w}}_{i,j}(t_1, t_2)$, or equivalently how rapidly the latent variable changes over time. The $\tau^{\mathrm{w}}_{i,j}$ parameters are estimated from the neural activity, together with the other DLAG parameters (see below). We follow the same conventions as in ref. [44], and fix $(\sigma^{\mathrm{w}}_{i,j})^2$ to a small value (10^{-3}). Note also that, under this definition, the process is normalized so that $k^{\mathrm{w}}_{i,j}(t_1, t_2) = 1$ for $t_1 = t_2$. Thus, the prior distribution of within-area latent variables $\mathbf{x}^{\mathrm{w}}_{i,:,t}$ in area i at each time t follows the standard normal distribution, $\mathcal{N}(0, I)$. This normalization removes model redundancy in the scaling of X^{w}_i and C^{w}_i.

Beyond describing within-area interactions, within-area variables are critical to the interpretability of across-area variables. As we will define below, across-area variables describe the activity of neurons in both areas. Within-area variables could, in principle, be formulated as a special case of across-area variables, where the loading coefficients to one area (the appropriate columns of C^{a}_1 or C^{a}_2 in equation (1)) are identically zero. If the model does not allow for within-area variables, then across-area variables must explain within-area activity in addition to across-area activity. Across-area variables could thus reflect a mixture of

within- and across-area activity in this case, obfuscating their interpretation as representing population activity patterns that are correlated across areas. The presence of within-area variables allows the across-area variables to isolate activity that is truly correlated across areas. This statistical phenomenon applies to other statistical models, and is not specific to DLAG[32,62]. See Supplementary Discussion for further mathematical discussion.

Across-area latent variables. We next describe across-area temporal structure. Across-area variables are different from within-area variables in two respects: (1) across-area variables are defined in pairs, where the elements of each pair correspond to the two areas, and (2) the elements of each pair are time delayed relative to each other (Fig. 2, centre column). Thus in contrast to our definition of within-area variables, in which we considered each area separately, we now consider across-area variables in both areas together: $\mathbf{x}^{\mathrm{a}}_{1,j,:} \in \mathbb{R}^T$ and $\mathbf{x}^{\mathrm{a}}_{2,j,:} \in \mathbb{R}^T$, the j^{th} rows of X^{a}_1 and X^{a}_2, respectively, for the j^{th} across-area variable.

The across-area latent variables of area 1 ($\mathbf{x}^{\mathrm{a}}_{1,j,:}$) and area 2 ($\mathbf{x}^{\mathrm{a}}_{2,j,:}$) belong to the same GP (Supplementary Fig. 1d). The $\mathbf{x}^{\mathrm{a}}_{1,j,:}$ are values of the GP sampled on a time grid. The $\mathbf{x}^{\mathrm{a}}_{2,j,:}$ are values of the same GP, also sampled on a time grid, but offset from the time grid of area 1 by a time delay. We define the GP for each across-area variable $j = 1, \ldots, p^{\mathrm{a}}$ as follows:

$$\begin{bmatrix} \mathbf{x}^{\mathrm{a}}_{1,j,:} \\ \mathbf{x}^{\mathrm{a}}_{2,j,:} \end{bmatrix} \sim \mathcal{N}\left(0, \begin{bmatrix} K^{\mathrm{a}}_{1,1,j} & K^{\mathrm{a}}_{1,2,j} \\ K^{\mathrm{a}}_{2,1,j} & K^{\mathrm{a}}_{2,2,j} \end{bmatrix}\right) \tag{6}$$

where $K^{\mathrm{a}}_{1,1,j} = K^{\mathrm{a}}_{2,2,j}$ describe the autocovariance of each across-area variable, and $K^{\mathrm{a}}_{1,2,j} = K^{\mathrm{a}\top}_{2,1,j}$ describe the cross-covariance that couples the two areas (Supplementary Fig. 1e).

To express the auto- and cross-covariance functions, we introduce additional notation. Specifically, we indicate brain areas with two subscripts, $i_1 = 1, 2$ and $i_2 = 1, 2$. Then, we define $K^{\mathrm{a}}_{i_1,i_2,j} \in \mathbb{R}^{T \times T}$ to be either the auto- or cross-covariance matrix between across-area variable $\mathbf{x}^{\mathrm{a}}_{i_1,j,:}$ in area i_1 and across-area variable $\mathbf{x}^{\mathrm{a}}_{i_2,j,:}$ in area i_2. We again choose to use the squared exponential function for GP covariances. Therefore, element (t_1, t_2) of each $K^{\mathrm{a}}_{i_1,i_2,j}$ can be computed as follows[45]:

$$k^{\mathrm{a}}_{i_1,i_2,j}(t_1, t_2) = \left(1 - (\sigma^{\mathrm{a}}_j)^2\right) \exp\left(-\frac{(\Delta t)^2}{2(\tau^{\mathrm{a}}_j)^2}\right) + (\sigma^{\mathrm{a}}_j)^2 \cdot \delta_{\Delta t} \tag{7}$$

$$\Delta t = \left(t_2 - D_{i_2,j}\right) - \left(t_1 - D_{i_1,j}\right) \tag{8}$$

where the characteristic timescale, $\tau^{\mathrm{a}}_j \in \mathbb{R}_{>0}$, and the GP noise variance, $(\sigma^{\mathrm{a}}_j)^2 \in (0, 1)$, are model parameters. $\delta_{\Delta t}$ is the Kronecker delta, which is 1 for $\Delta t = 0$ and 0 otherwise.

We also introduce two new parameters: the time delay to area i_1, $D_{i_1,j} \in \mathbb{R}$, and the time delay to area i_2, $D_{i_2,j} \in \mathbb{R}$. Notice that, when computing the autocovariance for area i (that is, $i_1 = i_2 = i$), the time delay parameters $D_{i_1,j}$ and $D_{i_2,j}$ are equal, and so Δt (equation (8)) reduces simply to the time difference $(t_2 - t_1)$, as in the within-area case (equation (5)). Time delays are therefore only relevant when computing the cross-covariance between area 1 and area 2. The time delay to area 1, $D_{1,j}$, and the time delay to area 2, $D_{2,j}$, by themselves have no physically meaningful interpretation. Their difference $D_j = D_{2,j} - D_{1,j}$, however, represents a well defined, continuous-valued time delay from area 1 to area 2. The sign of the relative time delay D_j indicates the directionality of the lead–lag relationship between areas captured by latent variable j (positive, area 1 leads area 2; negative, area 2 leads area 1), which we interpret as a description of inter-areal signal flow.

Both the characteristic timescales τ^{a}_j and relative delays D_j are estimated from the neural activity, together with the other DLAG parameters (Supplementary Note). More specifically, to ensure identifiability of time delay parameters, we designate area 1 as the reference area, and fix the delays for area 1 at 0, that is, $D_{1,j} = 0$ for all across-area variables $j = 1, \ldots, p^{\mathrm{a}}$. Then, each relative time delay D_j is simply $D_{2,j}$. Note that D_j need not be an integer multiple of the sampling period or spike count bin width of the neural activity. As in the within-area case, the across-area GP noise variance, $(\sigma^{\mathrm{a}}_j)^2$, is set to a small value (10^{-3}). Furthermore, the across-area GP is also normalized so that $k^{\mathrm{a}}_{i_1,i_2,j}(t_1, t_2) = 1$ if $\Delta t = 0$, thereby removing model redundancy in the scaling of X^{a}_i and C^{a}_i.

DLAG special cases. Finally, we consider some special cases of the DLAG model that illustrate its relationship to other dimensionality reduction methods. First, by fixing all time delays to zero ($D_j = 0$), and by removing within-area latent variables ($p^{\mathrm{w}}_1 = p^{\mathrm{w}}_2 = 0$), DLAG becomes equivalent to GPFA[44] applied to the two areas jointly. By removing instead the across-area latent variables ($p^{\mathrm{a}} = 0$), and keeping the within-area latent variables intact, DLAG becomes equivalent to GPFA applied to each area independently. Finally, by removing temporal smoothing (that is, in the limit as all GP noise parameters σ^{a}_j, $\sigma^{\mathrm{w}}_{i,j}$ approach 1), while retaining both within- and across-area latent variables, DLAG becomes similar to probabilistic CCA[42,43]. Whereas probabilistic CCA describes within-area activity via observation noise covariance matrices (R_i; see equation (47) of Supplementary Note), this special-case DLAG model would describe within-area activity via low-dimensional latent variables.

Selecting the numbers of within- and across-area latent variables. DLAG has three hyperparameters: p^{a}, the number of across-area latent variables; and p^{w}_1 and p^{w}_2, the number of within-area latent variables for each area. Model selection therefore poses a scaling challenge. Grid search over even a small range of within- and across-area dimensionalities can result in a large number of models that need to be fitted and validated. For example, considering just 10 possibilities for each type of latent variable would result in 1,000 candidate models. Thus, exhaustive search for the optimal DLAG model is impractical.

We therefore developed a streamlined cross-validation procedure that improves scalability. In brief, our model selection procedure occurs in two stages. First, we consider each area separately, and—using FA—we find the number of latent variables needed to explain the shared variance among neurons within each area. We reasoned that, while there is not a direct correspondence between the optimal number of latent variables in DLAG and FA models (because of temporal smoothing and other differences in model structure), it is unlikely that the total number of within- and across-area latent variables extracted by DLAG will exceed the FA dimensionality for an area (such a case would imply that there exists a neuron in, for example, area A that covaries with one or more neurons in area B, but no other neurons in area A). Hence we believe this approach to be reasonable given the computational benefits. We then use the FA dimensionality in each area to reduce the space of DLAG model candidates to a practical size.

In greater detail, we first applied FA to each area independently, and identified the optimal FA dimensionality through K-fold cross-validation (here we chose $K = 4$). We randomly split all trials into K equally sized partitions. For the k^{th} cross-validation fold ($k = 1, \ldots, K$), we held out the k^{th} partition of trials and fitted FA model parameters to the trials in the remaining $K - 1$ partitions. Using the fitted parameters, we evaluated the data log likelihood on the held-out trials. We repeated this procedure for each of the K folds and summed the held-out data log likelihoods computed for each fold. We refer to this value as the cross-validated data (log) likelihood. The FA model with the highest cross-validated data likelihood was taken as 'optimal'.

We then used the optimal FA dimensionalities (p^{FA}_i, $i = 1, 2$) to constrain the space of DLAG model candidates. In particular, we consider only DLAG models that satisfy $p^{\mathrm{a}} + p^{\mathrm{w}}_i = p^{\mathrm{FA}}_i$, for $i = 1, 2$; and $p^{\mathrm{a}} \leq \min(p^{\mathrm{FA}}_1, p^{\mathrm{FA}}_2)$. In words, we consider only DLAG models such that the number of within- and across-area latent variables in each area sum to that area's optimal FA dimensionality. Furthermore, the number of across-area latent variables is limited by the area with the smallest optimal FA dimensionality. Not only does this streamlined cross-validation approach provide an upper limit on the possible number of within- and across-area latent variables, it also effectively collapses the DLAG hyperparameter space from three free hyperparameters to one (across-area dimensionality, p^{a}), markedly improving scalability.

Among the model candidates within this constrained search range, we selected models that exhibited the largest cross-validated data likelihood, using the same K-fold cross-validation scheme as for FA. For each of the K folds, we evaluated (the log of) equation (31) of Supplementary Note on held-out trials using DLAG model parameters fitted to all remaining trials. We then took the cross-validated data log likelihood to be the sum (across the K folds) of held-out data log likelihoods. To further reduce runtime, we limited the number of expectation-maximization iterations during cross-validation to 1,000. The optimal DLAG model was then refitted to full convergence, where the data log likelihood improved from one iteration to the next by less than a preset tolerance (here we used 10^{-8}).

We also note that throughout this work we explicitly considered model candidates for which across-area dimensionality was zero ($p^{\mathrm{a}} = 0$): the two areas are independent, and any correlations between neurons are purely within-area. Similarly, we explicitly considered model candidates for which within-area dimensionalities were zero ($p^{\mathrm{w}}_1 = 0$ or $p^{\mathrm{w}}_2 = 0$): all variance shared among neurons in one area is attributed to their interactions with neurons in the other area. The case where all dimensionalities are zero ($p^{\mathrm{a}} = p^{\mathrm{w}}_1 = p^{\mathrm{w}}_2 = 0$) is equivalent to fitting a multivariate Gaussian distribution to the data with diagonal covariance (that is, all neurons are treated as independent). We similarly considered zero-dimensional FA models ($p^{\mathrm{FA}}_1 = 0$ or $p^{\mathrm{FA}}_2 = 0$) during the first stage of our model selection procedure, equivalent to fitting a multivariate Gaussian distribution with diagonal covariance to observations in the respective area. The inclusion of these zero-dimensionality model candidates protects against the identification of spurious interactions across or within areas.

Synthetic data generation. We generated synthetic datasets according to the DLAG generative model, so that we could leverage known ground truth to evaluate the accuracy of estimates and characterize DLAG's performance over a range of simulated conditions. We started by randomly generating the set of model parameters, θ (see equation (19) of Supplementary Note), subject to constraints informed by experimental data. For all datasets, we chose the numbers of neurons in each area on the basis of 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. Importantly, among these datasets, we included datasets without across- or within-area structure (that is, datasets for which across- or within-area dimensionality was zero), to test if our framework could identify such cases.

To ensure that synthetic datasets exhibited realistic noise levels, we first evaluated the strength of latent variables relative to the strength of single-neuron variability exhibited in the V1–V2 recordings. Specifically, we computed the 'signal-to-noise' ratio (where 'signal' is defined as the shared activity described by latent variables), $\mathrm{tr}(C_i C_i^{\top})/\mathrm{tr}(R_i)$, for V1 and V2 using the parameters of the optimal DLAG models fitted to each V1–V2 dataset. Representative values were 0.3 and 0.2 for V1 and V2, respectively. Then for each dataset, we generated our synthetic observation model parameters, C_i and R_i, as follows. We first drew the elements of C_i and a diagonal matrix $\Phi_i \in \mathbb{R}^{q_i \times q_i}$ from the standard normal distribution $\mathcal{N}(0, 1)$. Then, we set $R_i=\Phi_i\Phi_i^{\mathrm{T}}$ (so that R_i was a valid covariance matrix) and rescaled R_i such that area i exhibited the correct signal-to-noise ratio. The elements of the mean parameter $\mathbf{d}$ were also drawn from the standard normal distribution.

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 GP noise variances ($\{(\sigma_j^{a})^2\}_{j=1}^{p^{a}}$, $\{(\sigma_{1,j}^{w})^2\}_{j=1}^{p_1^{w}}$, $\{(\sigma_{2,j}^{w})^2\}_{j=1}^{p_2^{w}}$) were fixed at 10^{-3}. With all model parameters specified, we then generated $N=100$ independent and identically distributed trials ($\mathbf{x}_n$, $\mathbf{y}_n$, $n=1,\ldots,N$) according to equations (26) and (27) of Supplementary Note. Each trial comprised $T=50$ time points, corresponding to 1,000 ms sequences sampled with a period of 20 ms, to mimic the 20 ms spike count time bins used to analyse the experimental data.

Synthetic data performance metrics. To quantify DLAG's performance across all synthetic datasets, we employed a variety of metrics. We first consider the estimation of DLAG's observation model parameters. To assess the accuracy of loading matrix estimation (C_i^{a}, C_i^{w}; reported in Fig. 3 and Supplementary Figs. 2 and 7), we computed a normalized subspace error[66]:

$$e_{\mathrm{sub}} = \frac{\| (I - \hat{M}\left(\hat{M}^{\top}\hat{M}\right)^{-1}\hat{M}^{\top})M\|_F}{\| M\|_F} \tag{9}$$

where M is the appropriate ground truth parameter, $\hat{M}$ is the corresponding estimate and $\|\cdot\|_F$ is the Frobenius norm. e_{sub} quantifies the magnitude of the projection of the column space of M onto the null space of $\hat{M}$. A value of 1 indicates that the column space of M lies completely in the null space of $\hat{M}$, and therefore the estimate captures no component of the ground truth. A value of 0 indicates that the column space of $\hat{M}$ contains the full column space of M, and therefore the estimate captures all components of the ground truth. This metric offers two advantages: (1) it does not require that the columns of M and $\hat{M}$ are ordered in any way (the ordering of DLAG latent variables is arbitrary) and (2) it does not require that M and $\hat{M}$ have the same number of columns, so it can be used to compare the performance of models with different numbers of latent variables. We report the accuracy of loading matrix estimation as $1-e_{\mathrm{sub}}$ (Fig. 3). To assess the accuracy of estimating $\mathbf{d}$ and R (reported in Supplementary Figs. 2 and 7), we computed the normalized error

$$e_{\mathrm{vec}} = \frac{\| \mathbf{v} - \hat{\mathbf{v}}\|_2}{\| \mathbf{v}\|_2} \tag{10}$$

where $\mathbf{v}$ is either $\mathbf{d}$ or $\mathrm{diag}(R)$, and $\hat{\mathbf{v}}$ is the corresponding estimate.

We next consider the estimation of DLAG's state model parameters. Reporting the accuracy of delay and timescale estimates (Fig. 3 and Supplementary Figs. 2, 3 and 7) required explicitly matching estimated latent variables to the ground truth. Given the large number of synthetic datasets presented here, we automated this matching process as follows. First, for each area i, we took the unordered across- and within-area latent-variable estimates, $\hat{\mathbf{x}}_i^{a}$ and $\hat{\mathbf{x}}_i^{w}$, and computed the pairwise correlation between each estimated latent variable and each ground truth latent variable, $\mathbf{x}_i^{a}$ and $\mathbf{x}_i^{w}$, across all time points and trials. We then reordered the estimated latent variables to match the ground truth latent variables with which they showed the highest magnitude of correlation. To report delay and timescale estimation performance, we computed the absolute error between ground truth and (matched) estimated parameters, to express the error in units of time (ms).

Finally, we consider the moment-by-moment estimation of latent variables. As with the loading matrix, delay and timescale estimates, quantifying the accuracy of latent-variable estimates requires care since the sign and ordering of latent variables is arbitrary and will not, in general, match between estimates and the ground truth. First, let $\widetilde{\mathbf{x}}_i^{a} = [\mathbf{x}_{i,:,1}^{a\top} \cdots \mathbf{x}_{i,:,T}^{a\top}]^{\top} \in \mathbb{R}^{p^{a}T}$ be a collection of all (ground truth) across-area variables at all time points in area i. Similarly, let $\widetilde{\mathbf{x}}_i^{w} = [\mathbf{x}_{i,:,1}^{w\top} \cdots \mathbf{x}_{i,:,T}^{w\top}]^{\top} \in \mathbb{R}^{p_i^{w}T}$ be a collection of all (ground truth) within-area variables at all time points in area i. Finally, define $\widetilde{C}_i^{a} \in \mathbb{R}^{q_iT\times p^{a}T}$ and $\widetilde{C}_i^{w} \in \mathbb{R}^{q_iT\times p_i^{w}T}$ to be block diagonal matrices comprising T copies of the (ground truth) matrices C_i^{a} and C_i^{w}, respectively; and define $\widetilde{\mathbf{d}}_i \in \mathbb{R}^{q_iT}$ by vertically concatenating T copies of (the ground truth) $\mathbf{d}_i$. We'll denote the estimates of each of these values by $\hat{\widetilde{\mathbf{x}}}_i^{a}$, $\hat{\widetilde{\mathbf{x}}}_i^{w}$, $\hat{\widetilde{C}}_i^{a}$, $\hat{\widetilde{C}}_i^{w}$ and $\hat{\widetilde{\mathbf{d}}}_i$. The estimates $\hat{\widetilde{\mathbf{x}}}_i^{a}$ and $\hat{\widetilde{\mathbf{x}}}_i^{w}$ are posterior means, computed according to equation (30) of Supplementary Note.

Then, to separate the accuracy of across-area variable estimation from the accuracy of within-area variable estimation (as reported in Fig. 3 and Supplementary Fig. 2), we estimated denoised (smoothed) observations, using only across-area or only within-area latent-variable estimates:

$$\hat{\widetilde{\mathbf{y}}}_i^{*} = \hat{\widetilde{C}}_i^{*}\hat{\widetilde{\mathbf{x}}}_i^{*} + \hat{\widetilde{\mathbf{d}}}_i \tag{11}$$

where $\hat{\widetilde{\mathbf{y}}}_i^{*} = [\hat{\mathbf{y}}_{i,1}^{*\top} \cdots \hat{\mathbf{y}}_{i,T}^{*\top}]^{\top} \in \mathbb{R}^{q_iT}$. Here, the asterisk is used to indicate either 'a' or 'w' as a superscript, where observations have been denoised using only across- or within-area variable estimates, respectively. We then collect the denoised sequences on all N trials, $\hat{\widetilde{\mathbf{y}}}_{i,n}^{*}$, $n=1,\ldots,N$, into the matrix $\hat{Y}_i^{*} = [\hat{\widetilde{\mathbf{y}}}_{i,1}^{*} \cdots \hat{\widetilde{\mathbf{y}}}_{i,N}^{*}] \in \mathbb{R}^{q_iT\times N}$. Analogously, define $Y_i^{*} \in \mathbb{R}^{q_iT\times N}$ to be the set of ground truth sequences generated before adding noise (that is, the noise term $\boldsymbol{\varepsilon}_i$, defined in equation (2)).

We then computed the R^2 value between estimated and (noiseless) ground truth sequences:

$$R^2 = 1 - \frac{\| Y_i^{*} - \hat{Y}_i^{*}\|_F^2}{\| Y_i^{*} - \bar{Y}_i^{*}\|_F^2} \tag{12}$$

where $\bar{Y}_i^{*} = [\bar{\mathbf{y}}_i^{*} \cdots \bar{\mathbf{y}}_i^{*}] \in \mathbb{R}^{q_iT\times N}$ is constructed by horizontally concatenating N copies of the sample mean for each neuron in the ground truth Y_i^{*}, taken over all time points and trials ($\bar{\mathbf{y}}_i^{*} \in \mathbb{R}^{q_iT}$). Note that, in the multivariate case, $R^2 \in (-\infty, 1]$, where a negative value implies that estimates predict the ground truth less accurately than simply the sample mean.

Visual stimuli and neural recordings. Animal procedures and recording details have been described in previous work[26,67]. Briefly, animals (*Macaca fascicularis*, young adult males) were anaesthetized with ketamine (10 mg kg^{-1}) and maintained on isoflurane (1–2%) during surgery. Recordings were performed under sufentanil (typically 6–18 μg kg^{-1} h^{-1}) anaesthesia. Vecuronium bromide (150 μg kg^{-1} h^{-1}) was used to prevent eye movements. The duration of each experiment (which comprised multiple recording sessions) varied from 5 to 7 d. All procedures were approved by the Institutional Animal Care and Use Committee of the Albert Einstein College of Medicine.

The data analysed here are those reported in refs. [36,41], and a subset of recording sessions reported in ref. [26]. Activity in V1 output layers was recorded using a 96-channel Utah array (400 μm interelectrode spacing, 1 mm length, inserted to a nominal depth of 600 μm). We recorded V2 activity using a set of electrodes/tetrodes (interelectrode spacing 300 μm) whose depth could be controlled independently (Thomas Recording). These electrodes were lowered through V1 and the underlying white matter, and then into V2. Within V2, we targeted neurons in the input layers. We verified that the recordings were performed in the input layers using measurements of the depth in V2 cortex, histological confirmation (in a subset of recordings) and correlation measurements. For complete details see refs. [67] and [26]. Voltage snippets that exceeded a user-defined threshold were digitized and sorted offline. The sampled neurons had spatial receptive fields within 2–4° of the fovea, in the lower visual field.

We measured responses evoked by drifting sinusoidal gratings (1–1.1 cycles per degree; drift rate 6.25 Hz; 2.6–4.95° in diameter; full contrast, defined as Michelson contrast, $(L_{\max}-L_{\min})/(L_{\max}+L_{\min})$, where $L_{\min}$ is 0 cd m^{-2} and $L_{\max}$ is 80 cd m^{-2}) at eight different orientations (22.5° steps), on a calibrated cathode-ray tube monitor placed 110 cm from the animal (1,024×768 pixel resolution at a 100 Hz refresh rate; Expo: http://sites.google.com/a/nyu.edu/expo). Each stimulus was presented 400 times for 1.28 s. Each presentation was preceded by an interstimulus interval of 1.5 s during which a grey screen was presented.

We recorded neuronal activity in three animals. In two of the animals, we recorded in two different but nearby locations in V2, providing distinct middle-layer populations, yielding a total of five recording sessions. We treated responses to each of the eight stimuli in each session separately, yielding a total of 40 'datasets'.

Data preprocessing. We counted spikes in 20 ms time bins during the 1.28 s stimulus presentation period (64 bins per trial). For all analyses corresponding to each recording session, we excluded neurons that fired fewer than 0.5 spikes s^{-1}, on average, across all trials and all grating orientations. Because we were interested in V1–V2 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[68]. We then applied DLAG to each dataset separately.

Intra-areal and subsampled population comparisons. To contrast with the V1–V2 results, we also used DLAG to characterize the interactions between two

V1 subpopulations. For each dataset, we randomly split V1 into two equally sized subpopulations (for datasets with an odd number of V1 neurons, we discarded one neuron at random). Each subpopulation was labelled arbitrarily as either 'V1a' or 'V1b' (Fig. 4c). We then applied DLAG to dissect these V1a–V1b interactions in a manner identical to V1–V2 (Figs. 5 and 6).

We also sought to understand the extent to which the V1–V2 results were driven by disparities in population size between V1 and V2 (Supplementary Fig. 9). For each dataset, we therefore randomly subsampled the V1 population to match the size of the V2 population. We then applied DLAG to each subsampled dataset in the same manner as above.

Variance explained by DLAG latent variables. After fitting a DLAG model to each experimental dataset, we sought to compare the relative strengths of across- or within-area latent variables extracted from the same dataset (as in Fig. 5) and across different datasets (as in Fig. 6b). To quantify these comparisons, we computed the variance explained by each latent variable, as derived from fitted model parameters. From equation (1), the total variance in area i simplifies to

$$\mathrm{var}_{\mathrm{total}} = \mathrm{tr}\left(C_i^{\mathrm{a}} C_i^{\mathrm{a}\top} + C_i^{\mathrm{w}} C_i^{\mathrm{w}\top} + R_i\right) \tag{13}$$

By inspection, the total variance decomposes into three separable components: $\mathrm{tr}(C_i^{\mathrm{a}} C_i^{\mathrm{a}\top})$, the variance due to across-area activity; $\mathrm{tr}(C_i^{\mathrm{w}} C_i^{\mathrm{w}\top})$, the variance due to within-area activity, and $\mathrm{tr}(R_i)$, the variance that is independent to each neuron. In fact, the across-area and within-area components can be decomposed further into contributions by individual latent variables. Let $\mathbf{c}_{i,j}^{\mathrm{a}} \in \mathbb{R}^{q_i}$ be the j^{th} column of C_i^{a}, and $\mathbf{c}_{i,j}^{\mathrm{w}} \in \mathbb{R}^{q_i}$ be the j^{th} column of C_i^{w}. Then, $\mathrm{tr}(C_i^{\mathrm{a}} C_i^{\mathrm{a}\top}) = \sum_{j=1}^{p^{\mathrm{a}}} \| \mathbf{c}_{i,j}^{\mathrm{a}} \|_2^2$, and $\mathrm{tr}(C_i^{\mathrm{w}} C_i^{\mathrm{w}\top}) = \sum_{j=1}^{p_i^{\mathrm{w}}} \| \mathbf{c}_{i,j}^{\mathrm{w}} \|_2^2$.

Because we were interested in variance shared among neurons, rather than independent to each neuron, we focused on the variance components involving C_i^{a} and C_i^{w}, rather than R_i. Furthermore, since the total variance of recorded neural activity may vary widely across animals, stimuli and recording sessions, we computed two normalized metrics to facilitate comparison of these shared variance components across datasets. First, let $\mathbf{c}_{i,j}$ be the j^{th} column of C_i, where $C_i = [C_i^{\mathrm{a}}\ C_i^{\mathrm{w}}]$ is the same as in equation (22) of Supplementary Note. To visualize the relative strength of latent variables in each area (Fig. 5), we computed

$$\alpha_{i,j} = \frac{\| \mathbf{c}_{i,j} \|_2^2}{\mathrm{tr}\left(C_i^{\mathrm{a}} C_i^{\mathrm{a}\top} + C_i^{\mathrm{w}} C_i^{\mathrm{w}\top}\right)} \tag{14}$$

that is, the fraction of shared variance explained by latent variable j in area i. We then displayed latent time courses multiplied by the appropriate $\alpha_{i,j}$ at each time point. Similarly, to quantify the strength of across-area activity (relative to within-area activity) in each area (Fig. 6b), we computed

$$\alpha_i^{\mathrm{a}} = \frac{\mathrm{tr}\left(C_i^{\mathrm{a}} C_i^{\mathrm{a}\top}\right)}{\mathrm{tr}\left(C_i^{\mathrm{a}} C_i^{\mathrm{a}\top} + C_i^{\mathrm{w}} C_i^{\mathrm{w}\top}\right)} \tag{15}$$

that is, the fraction of shared variance explained by all across-area latent variables in area i.

Uncertainty of estimated delays. DLAG's performance on the synthetic data presented here suggests that time delays are estimated with high accuracy and precision. For our neural recordings, however, where no 'ground truth' is accessible, we sought to assess the certainty with which fitted delay parameters were indeed positive or negative—indicating a particular direction of inter-areal signal flow. We therefore developed the following non-parametric bootstrap procedure.

First, consider a DLAG model that has been fitted to a particular dataset with N trials. We construct a bootstrap sample $b = 1, \ldots, B$ from this dataset by selecting N trials uniformly at random with replacement (here we used $B = 1{,}000$). Then, let ℓ_b be the data log likelihood of the DLAG model evaluated on bootstrap sample b, and let $\ell_{b,j=0}$ be the data log likelihood of the same DLAG model evaluated on bootstrap sample b, but for which D_j, the delay for across-area latent variable j, has been set to zero (all other model parameters remain unaltered).

To compare the performance of this 'zero-delay' model with the performance of the original model, we define the following statistic:

$$\Delta\ell_{b,j=0} = \ell_b - \ell_{b,j=0}. \tag{16}$$

If the zero-delay model performed at least as well as the original DLAG model (equivalently, $\Delta\ell_{b,j=0} \leq 0$) on 5% or more of the bootstrap samples, then we could not say, with sufficient certainty, that the delay for across-area variable j was strictly positive or strictly negative. Otherwise, we took the magnitude of the delay for across-area variable j to differ significantly from zero.

For each of our V1–V2 datasets, then, this procedure allowed us to label some delays as ambiguous, where the corresponding population signal could not be confidently categorized as flowing in one direction or the other (Fig. 6c). Finally, note that the concept of ambiguity defined here is distinct from the concept of a variable's importance in describing observed neural activity: for example, an across-area variable with an ambiguous time delay between areas could, in principle, still explain a large portion of an area's shared variance.

Across-area prediction. As described above, we selected the number of within- and across-area latent variables for DLAG models using cross-validated data log likelihood (from equation (31) of Supplementary Note). Cross-validated data log likelihood offers a principled performance metric, as it is precisely the (training) data log likelihood that a fitted DLAG model maximizes, and it fits within DLAG's probabilistic framework. However, interpretation of the performance differences between models can be difficult given the scale of log likelihood values. Furthermore, log likelihood values can vary markedly from dataset to dataset, often by orders of magnitude. We therefore sought an alternative metric that facilitates more intuitive comparison between models/methods (Supplementary Figs. 12 and 15) and across datasets (Supplementary Fig. 11).

Towards this end, we developed a leave-group-out prediction procedure that measures a model's ability to capture interactions across areas (similar to the leave-neuron-out prediction, or 'co-smoothing', procedures in refs. [44,69]). Our goal, therefore, is to use a fitted model to predict the unobserved activity of held-out neurons in one area, given the observed activity of neurons in the other area. Let us first collect observed variables (for one trial) in a manner that highlights group structure. 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 $t = 1, \ldots, T$.

To predict $\widetilde{\mathbf{y}}_2$ from $\widetilde{\mathbf{y}}_1$, we use the conditional distribution of $\widetilde{\mathbf{y}}_2$ given $\widetilde{\mathbf{y}}_1$, $P(\widetilde{\mathbf{y}}_2 | \widetilde{\mathbf{y}}_1)$, which can be obtained from the joint distribution, $P(\widetilde{\mathbf{y}}_1, \widetilde{\mathbf{y}}_2)$. For a derivation and discussion of $P(\widetilde{\mathbf{y}}_1, \widetilde{\mathbf{y}}_2)$, see Supplementary Discussion (equation (52)). From $P(\widetilde{\mathbf{y}}_2 | \widetilde{\mathbf{y}}_1)$, we take predictions to be the expected value of activity in area 2 given activity in area 1:

$$\hat{\widetilde{\mathbf{y}}}_2 = \mathbb{E}\left[\widetilde{\mathbf{y}}_2 \,|\, \widetilde{\mathbf{y}}_1\right] = \widetilde{C}_2^{\mathrm{a}} \widetilde{K}_{2,1}^{\mathrm{a}} \widetilde{C}_1^{\mathrm{a}\top} \left(\widetilde{C}_1^{\mathrm{a}} \widetilde{K}_{1,1}^{\mathrm{a}} \widetilde{C}_1^{\mathrm{a}\top} + \widetilde{C}_1^{\mathrm{w}} \widetilde{K}_1^{\mathrm{w}} \widetilde{C}_1^{\mathrm{w}\top} + \widetilde{R}_1\right)^{-1} \left(\widetilde{\mathbf{y}}_1 - \widetilde{\mathbf{d}}_1\right) + \widetilde{\mathbf{d}}_2 \tag{17}$$

where $\widetilde{C}_1^{\mathrm{a}} \in \mathbb{R}^{q_1 T \times p^{\mathrm{a}} T}$, $\widetilde{C}_1^{\mathrm{w}} \in \mathbb{R}^{q_1 T \times p_1^{\mathrm{w}} T}$, $\widetilde{C}_2^{\mathrm{a}} \in \mathbb{R}^{q_2 T \times p^{\mathrm{a}} T}$, $\widetilde{C}_2^{\mathrm{w}} \in \mathbb{R}^{q_2 T \times p_2^{\mathrm{w}} T}$, $\widetilde{R}_1 \in \mathbb{S}^{q_1 T \times q_1 T}$ 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. The GP covariance matrices $\widetilde{K}_1^{\mathrm{w}} \in \mathbb{S}^{p_1^{\mathrm{w}} T \times p_1^{\mathrm{w}} T}$, $\widetilde{K}_{1,1}^{\mathrm{a}} \in \mathbb{R}^{p^{\mathrm{a}} T \times p^{\mathrm{a}} T}$ and $\widetilde{K}_{2,1}^{\mathrm{a}} \in \mathbb{R}^{p^{\mathrm{a}} T \times p^{\mathrm{a}} T}$ are defined in equations (50) and (51) of Supplementary Discussion. We similarly predict $\widetilde{\mathbf{y}}_1$ from $\widetilde{\mathbf{y}}_2$ using $\mathbb{E}\left[\widetilde{\mathbf{y}}_1 | \widetilde{\mathbf{y}}_2\right]$.

We next use equation (17) to define a cross-validated measure of a model's across-area predictive performance. Assume that we are given the parameters of a DLAG model fit to training data (equation (19) of Supplementary Note). Then let $\widetilde{\mathbf{y}}_{i,n}$ be the activity of area i on trial n of a held-out validation set, and let $\hat{\widetilde{\mathbf{y}}}_{i,n}$ be its predicted value given by equation (17). Collect these values across all $n = 1, \ldots, N$ held-out validation set trials into the respective matrices $Y_i = [\widetilde{\mathbf{y}}_{i,1} \cdots \widetilde{\mathbf{y}}_{i,N}] \in \mathbb{R}^{q_i T \times N}$ and $\hat{Y}_i = [\hat{\widetilde{\mathbf{y}}}_{i,1} \cdots \hat{\widetilde{\mathbf{y}}}_{i,N}] \in \mathbb{R}^{q_i T \times N}$. We then define a leave-group-out R^2 value as follows:

$$R^2_{\mathrm{lgo}} = 1 - \frac{\| Y_1 - \hat{Y}_1 \|_{\mathrm{F}}^2 + \| Y_2 - \hat{Y}_2 \|_{\mathrm{F}}^2}{\| Y_1 - \bar{Y}_1 \|_{\mathrm{F}}^2 + \| Y_2 - \bar{Y}_2 \|_{\mathrm{F}}^2} \tag{18}$$

where $\bar{Y}_i = [\bar{\mathbf{y}}_i \cdots \bar{\mathbf{y}}_i] \in \mathbb{R}^{q_i T \times N}$ is constructed by horizontally concatenating N copies of the sample mean for each neuron in observations Y_i, taken over all time points and trials ($\bar{\mathbf{y}}_i \in \mathbb{R}^{q_i T}$). In K-fold cross-validation, we evaluate R^2_{lgo} on each of the K validation sets, and report the average value over all K.

In a typical multivariate regression setting, R^2 is an asymmetric measure of predictive performance: prediction of $\widetilde{\mathbf{y}}_2$ from $\widetilde{\mathbf{y}}_1$ yields a different R^2 value than does prediction of $\widetilde{\mathbf{y}}_1$ from $\widetilde{\mathbf{y}}_2$. In contrast, R^2_{lgo} is a symmetric measure that aggregates predictions in both directions. Like R^2, $R^2_{\mathrm{lgo}} \in (-\infty, 1]$, where a value of 1 implies perfect prediction of neural activity, and a negative value implies that estimates predict neural activity less accurately than simply the sample mean. R^2_{lgo} is normalized by the total variance of neural activity within each dataset, thereby facilitating comparison across datasets, in which the variance of neural activity could vary widely. This more intuitive comparison across datasets (compared with log likelihood) comes at the expense of a principled characterization of performance within DLAG's probabilistic framework, and we emphasize that across-area prediction is not the objective that a fitted DLAG model is designed to maximize.

Statistics and reproducibility. All statistical analyses described here were carried out in MATLAB (MathWorks). To assess whether across-area strength was significantly greater in V2 than in V1, we performed a one-sided paired sign test (Fig. 6b). To assess whether V1–V2 time delays were significantly less than zero, we performed a one-sided one-sample sign test (Fig. 6c). To assess whether

within-V2 GP timescales were significantly longer than within-V1 GP timescales, we performed a one-sided Wilcoxon rank sum test (Fig. 6d). *P* values are reported in the caption of Fig. 6.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

V1–V2 data are available at the CRCNS data sharing website at https://doi.org/10.6080/K0B27SHN (ref. [70]). Naturalistic texture images are available on the Multiband Texture Database at http://multibandtexture.recherche.usherbrooke.ca/original_brodatz.html and on the Salzburg Texture Image Database at https://wavelab.at/sources/STex. Source Data for Figs. 3–6 are available for this Article.

Code availability

A MATLAB implementation of DLAG is available on GitHub at https://github.com/egokcen/DLAG and on Zenodo at https://doi.org/10.5281/zenodo.6654831 (ref. [71]).

Received: 16 October 2021; Accepted: 21 June 2022;
Published online: 18 August 2022

References

1. Ahrens, M. B. et al. Brain-wide neuronal dynamics during motor adaptation in zebrafish. *Nature* **485**, 471–477 (2012).
2. Yang, W. & Yuste, R. In vivo imaging of neural activity. *Nat. Methods* **14**, 349–359 (2017).
3. Jun, J. J. et al. Fully integrated silicon probes for high-density recording of neural activity. *Nature* **551**, 232–236 (2017).
4. Steinmetz, N. A., Zatka-Haas, P., Carandini, M. & Harris, K. D. Distributed coding of choice, action and engagement across the mouse brain. *Nature* **576**, 266–273 (2019).
5. Kohn, A. et al. Principles of corticocortical communication: proposed schemes and design considerations. *Trends Neurosci.* **43**, 725–737 (2020).
6. Lamme, V. A., Supèr, H. & Spekreijse, H. Feedforward, horizontal, and feedback processing in the visual cortex. *Curr. Opin. Neurobiol.* **8**, 529–535 (1998).
7. Angelucci, A. & Bressloff, P. C. Contribution of feedforward, lateral and feedback connections to the classical receptive field center and extra-classical receptive field surround of primate V1 neurons. *Prog. Brain Res.* **154**, 93–120 (2006).
8. Gilbert, C. D. & Li, W. Top-down influences on visual processing. *Nat. Rev. Neurosci.* **14**, 350–363 (2013).
9. Harris, K. D. & Mrsic-Flogel, T. D. Cortical connectivity and sensory coding. *Nature* **503**, 51–58 (2013).
10. Miller, E. K., Lundqvist, M. & Bastos, A. M. Working Memory 2.0. *Neuron* **100**, 463–475 (2018).
11. Shadmehr, R. & Krakauer, J. W. A computational neuroanatomy for motor control. *Exp. Brain Res.* **185**, 359–381 (2008).
12. Keemink, S. W. & Machens, C. K. Decoding and encoding (de)mixed population responses. *Curr. Opin. Neurobiol.* **58**, 112–121 (2019).
13. Schmolesky, M. T. et al. Signal timing across the macaque visual system. *J. Neurophysiol.* **79**, 3272–3278 (1998).
14. Hernández, A. et al. Decoding a perceptual decision process across cortex. *Neuron* **66**, 300–314 (2010).
15. Siegel, M., Buschman, T. J. & Miller, E. K. Cortical information flow during flexible sensorimotor decisions. *Science* **348**, 1352–1355 (2015).
16. Supèr, H., Spekreijse, H. & Lamme, V. A. F. Two distinct modes of sensory processing observed in monkey primary visual cortex (V1). *Nat. Neurosci.* **4**, 304–310 (2001).
17. Pooresmaeili, A., Poort, J. & Roelfsema, P. R. Simultaneous selection by object-based attention in visual and frontal cortex. *Proc. Natl Acad. Sci. USA* **111**, 6467–6472 (2014).
18. Chen, M. et al. Incremental integration of global contours through interplay between visual cortical areas. *Neuron* **82**, 682–694 (2014).
19. Schwiedrzik, C. M. & Freiwald, W. A. High-level prediction signals in a low-level area of the macaque face-processing hierarchy. *Neuron* **96**, 89–97 (2017).
20. Issa, E. B., Cadieu, C. F. & DiCarlo, J. J. Neural dynamics at successive stages of the ventral visual stream are consistent with hierarchical error signals. *eLife* **7**, e42870 (2018).
21. Reid, R. C. & Alonso, J. M. Specificity of monosynaptic connections from thalamus to visual cortex. *Nature* **378**, 281–284 (1995).
22. Roe, A. W. & Ts'o, D. Y. Specificity of color connectivity between primate V1 and V2. *J. Neurophysiol.* **82**, 2719–2730 (1999).
23. Nowak, L. G., Munk, M., James, A. C., Girard, P. & Bullier, J. Cross-correlation study of the temporal interactions between areas V1 and V2 of the macaque monkey. *J. Neurophysiol.* **81**, 1057–1074 (1999).
24. Jia, X., Tanabe, S. & Kohn, A. Gamma and the coordination of spiking activity in early visual cortex. *Neuron* **77**, 762–774 (2013).
25. Oemisch, M., Westendorff, S., Everling, S. & Womelsdorf, T. Interareal spike-train correlations of anterior cingulate and dorsal prefrontal cortex during attention shifts. *J. Neurosci.* **35**, 13076–13089 (2015).
26. Zandvakili, A. & Kohn, A. Coordinated neuronal activity enhances corticocortical communication. *Neuron* **87**, 827–839 (2015).
27. Campo, A. T. et al. Feed-forward information and zero-lag synchronization in the sensory thalamocortical circuit are modulated during stimulus perception. *Proc. Natl Acad. Sci. USA* **116**, 7513–7522 (2019).
28. Gregoriou, G. G., Gotts, S. J., Zhou, H. & Desimone, R. High-frequency, long-range coupling between prefrontal and visual cortex during attention. *Science* **324**, 1207–1210 (2009).
29. Salazar, R. F., Dotson, N. M., Bressler, S. L. & Gray, C. M. Content-specific fronto-parietal synchronization during visual working memory. *Science* **338**, 1097–1100 (2012).
30. van Kerkoerle, T. et al. Alpha and gamma oscillations characterize feedback and feedforward processing in monkey visual cortex. *Proc. Natl Acad. Sci. USA* **111**, 14332–14341 (2014).
31. Bastos, A. M., Vezoli, J. & Fries, P. Communication through coherence with inter-areal delays. *Curr. Opin. Neurobiol.* **31**, 173–180 (2015).
32. Semedo, J. D., Gokcen, E., Machens, C. K., Kohn, A. & Yu, B. M. Statistical methods for dissecting interactions between brain areas. *Curr. Opin. Neurobiol.* **65**, 59–69 (2020).
33. Kang, B. & Druckmann, S. Approaches to inferring multi-regional interactions from simultaneous population recordings. *Curr. Opin. Neurobiol.* **65**, 108–119 (2020).
34. Keeley, S. L., Zoltowski, D. M., Aoi, M. C. & Pillow, J. W. Modeling statistical dependencies in multi-region spike train data. *Curr. Opin. Neurobiol.* **65**, 194–202 (2020).
35. Kaufman, M. T., Churchland, M. M., Ryu, S. I. & Shenoy, K. V. Cortical activity in the null space: permitting preparation without movement. *Nat. Neurosci.* **17**, 440–448 (2014).
36. Semedo, J. D., Zandvakili, A., Machens, C. K., Yu, B. M. & Kohn, A. Cortical areas interact through a communication subspace. *Neuron* **102**, 249–259 (2019).
37. Perich, M. G., Gallego, J. A. & Miller, L. E. A neural population mechanism for rapid learning. *Neuron* **100**, 964–976 (2018).
38. Srinath, R., Ruff, D. A. & Cohen, M. R. Attention improves information flow between neuronal populations without changing the communication subspace. *Curr. Biol.* **31**, 5299–5313 (2021).
39. Veuthey, T. L., Derosier, K., Kondapavulur, S. & Ganguly, K. Single-trial cross-area neural population dynamics during long-term skill learning. *Nat. Commun.* **11**, 4057 (2020).
40. Chen, G., Kang, B., Lindsey, J., Druckmann, S. & Li, N. Modularity and robustness of frontal cortical networks. *Cell* **184**, 3717–3730 (2021).
41. Semedo, J. D. et al. Feedforward and feedback interactions between visual cortical areas use different population activity patterns. *Nat. Commun.* **13**, 1099 (2022).
42. Bach, F. R. & Jordan, M. I. *A Probabilistic Interpretation of Canonical Correlation Analysis* Technical Report 688 (Department of Statistics, University of California, Berkeley, 2005).
43. Archambeau, C. & Bach, F. Sparse probabilistic projections. *Adv. Neural Inf. Process. Syst.* **21**, 73–80 (2008).
44. Yu, B. M. et al. Gaussian-process factor analysis for low-dimensional single-trial analysis of neural population activity. *J. Neurophysiol.* **102**, 614–635 (2009).
45. Lakshmanan, K. C., Sadtler, P. T., Tyler-Kabara, E. C., Batista, A. P. & Yu, B. M. Extracting low-dimensional latent structure from time series in the presence of delays. *Neural Comput.* **27**, 1825–1856 (2015).
46. Felleman, D. J. & Van Essen, D. C. Distributed hierarchical processing in the primate cerebral cortex. *Cereb. Cortex* **1**, 1–47 (1991).
47. Markov, N. T. et al. Cortical high-density counterstream architectures. *Science* **342**, 1238406 (2013).
48. Smith, M. A., Kohn, A. & Movshon, J. A. Glass pattern responses in macaque V2 neurons. *J. Vision* **7**, 5 (2007).
49. Murray, J. D. et al. A hierarchy of intrinsic timescales across primate cortex. *Nat. Neurosci.* **17**, 1661–1663 (2014).
50. Runyan, C. A., Piasini, E., Panzeri, S. & Harvey, C. D. Distinct timescales of population coding across cortex. *Nature* **548**, 92–96 (2017).
51. Siegle, J. H. et al. Survey of spiking in the mouse visual system reveals functional hierarchy. *Nature* **592**, 86–92 (2021).
52. Zhuang, X., Yang, Z. & Cordes, D. A technical review of canonical correlation analysis for neuroscience applications. *Hum. Brain Mapp.* **41**, 3807–3833 (2020).
53. Kamiński, M., Ding, M., Truccolo, W. A. & Bressler, S. L. Evaluating causal relations in neural systems: Granger causality, directed transfer function and statistical assessment of significance. *Biol. Cybern.* **85**, 145–157 (2001).

54. Quinn, C. J., Coleman, T. P., Kiyavash, N. & Hatsopoulos, N. G. Estimating the directed information to infer causal relationships in ensemble neural spike train recordings. *J. Comput. Neurosci.* **30**, 17–44 (2011).
55. Kim, S., Putrino, D., Ghosh, S. & Brown, E. N. A Granger causality measure for point process models of ensemble neural spiking activity. *PLoS Comput. Biol.* **7**, e1001110 (2011).
56. Pillow, J. W. et al. Spatio-temporal correlations and visual signalling in a complete neuronal population. *Nature* **454**, 995–999 (2008).
57. Truccolo, W., Hochberg, L. R. & Donoghue, J. P. Collective dynamics in human and monkey sensorimotor cortex: predicting single neuron spikes. *Nat. Neurosci.* **13**, 105–111 (2010).
58. Perich, M. G. et al. Inferring brain-wide interactions using data-constrained recurrent neural network models. Preprint at https://doi.org/10.1101/2020.12.18.423348 (2021).
59. Rodu, J., Klein, N., Brincat, S. L., Miller, E. K. & Kass, R. E. Detecting multivariate cross-correlation between brain regions. *J. Neurophysiol.* **120**, 1962–1972 (2018).
60. Bong, H. et al. Latent dynamic factor analysis of high-dimensional neural recordings. *Adv. Neural Inf. Process. Syst.* **33**, 16446–16456 (2020).
61. Keeley, S., Aoi, M., Yu, Y., Smith, S. & Pillow, J. W. Identifying signal and noise structure in neural population activity with Gaussian process factor models. *Adv. Neural Inf. Process. Syst.* **33**, 13795–13805 (2020).
62. Semedo, J., Zandvakili, A., Kohn, A., Machens, C. K. & Yu, B. M. Extracting latent structure from multiple interacting neural populations. *Adv. Neural Inf. Process. Syst.* **27**, 2942–2950 (2014).
63. Glaser, J., Whiteway, M., Cunningham, J. P., Paninski, L. & Linderman, S. Recurrent switching dynamical systems models for multiple interacting neural populations. *Adv. Neural Inf. Process. Syst.* **33**, 14867–14878 (2020).
64. Reid, A. T. et al. Advancing functional connectivity research from association to causation. *Nat. Neurosci.* **22**, 1751–1760 (2019).
65. Rasmussen, C. E. & Williams, C. K. I. *Gaussian Processes for Machine Learning* (MIT Press, 2006).
66. Golub, G. H. & Van Loan, C. F. *Matrix Computations* 4th edn (Johns Hopkins Univ. Press, 2013).
67. Smith, M. A. & Kohn, A. Spatial and temporal scales of neuronal correlation in primary visual cortex. *J. Neurosci.* **28**, 12591–12603 (2008).
68. Cowley, B. R. et al. Slow drift of neural activity as a signature of impulsivity in macaque visual and prefrontal cortex. *Neuron* **108**, 551–567 (2020).
69. Pei, F. et al. Neural Latents Benchmark '21: evaluating latent variable models of neural population activity. Preprint at https://doi.org/10.48550/arXiv.2109.04463 (2022).
70. Zandvakili, A. & Kohn, A. Simultaneous V1–V2 neuronal population recordings in anesthetized macaque monkeys. *CRCNS* https://doi.org/10.6080/K0B27SHN (2019).
71. Gokcen, E. egokcen/DLAG: v1.0.0. *Zenodo* https://doi.org/10.5281/zenodo.6654831 (2022).

Acknowledgements

This work was supported by the Dowd Fellowship (E.G.), Simons Collaboration on the Global Brain 542999 (A.K.), 543009 (C.K.M.), 543065 (B.M.Y.), 364994 (A.K., B.M.Y.), NIH R01 EY028626 (A.K.), NIH U01 NS094288 (C.K.M.), NIH R01 HD071686 (B.M.Y.), NIH CRCNS R01 NS105318 (B.M.Y.), NSF NCS BCS 1533672 and 1734916 (B.M.Y.), NIH CRCNS R01 MH118929 (B.M.Y.) and NIH R01 EB026953 (B.M.Y.).

Author contributions

E.G., A.I.J., J.D.S., A.K., C.K.M. and B.M.Y. designed the analyses. E.G. derived and implemented DLAG, and performed all analyses. A.I.J., A.Z. and A.K. designed and performed the experiments. E.G., A.I.J., A.K., C.K.M. and B.M.Y. wrote the manuscript. All authors discussed the results and commented on the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s43588-022-00282-5.

Correspondence and requests for materials should be addressed to Byron M. Yu.

Peer review information *Nature Computational Science* thanks Matthew Kaufman, Stephen Keeley, Stefano Recanatesi and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Primary Handling Editor: Ananya Rastogi, in collaboration with the *Nature Computational Science* team.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.