# State-dependent modulation of dependencies in a laminar cortical circuit

Anirban Das[2,3,4,#], Alec G. Sheffield[1,#], Anirvan S. Nandy[1,3,†] and Monika P. Jadi[1,2,3,†,*]

[1]Interdepartmental Neuroscience Program, Yale University, New Haven, CT 06511

[2]Department of Psychiatry, Yale University, New Haven, CT 06511

[3]Department of Neuroscience, Yale University, New Haven, CT 06511

[4]Design and Patterning AI Group, Intel Corp., Hillsboro, Oregon 97124

[#]equal contribution

[†]senior author

[*]corresponding author (monika.jadi@yale.edu)

## ABSTRACT

Adaptive information processing, comprised of local computations and their efficient routing, is crucial for flexible brain function. Spatial attention is a quintessential example of this adaptive process. It is critical for recognizing and interacting with behaviorally relevant objects in a cluttered environment[1]. Object recognition is mediated by ensembles of computational units distributed across the ventral visual hierarchy. How the deployment of spatial attention aids these hierarchical computations is unclear. Based on pairwise correlation analysis, two key mechanisms have been proposed: First is an improvement in the efficacy of unique information directed from one encoding stage to another, suggested by evidence along the visual hierarchy[2-6]. Based on the theoretical results that even weak correlated variability can substantially limit the encoding capacity of a neuronal pool[7], a second proposal is an improvement in the sensory information capacity of an encoding stage through a reduction in shared fluctuations[8,9]. However, pairwise analyses capture both unique and shared components of fluctuations, and therefore cannot disambiguate the proposed mechanisms. To test these proposals, it is crucial to estimate the

attentional modulation of unique information flow across and shared information within the stages of the visual hierarchy. We investigated this in the multi-stage laminar network of visual area V4, an area strongly modulated by attention[10-12]. Using network-based statistical modeling, we estimated the strength of inter-layer information flow by measuring statistical dependencies that reflect how the cortical layers *uniquely* drive each other's neural activity. We quantified their modulation across attention conditions (attend-in vs. attend-away) in a change detection task and found that deployment of attention indeed strengthened unique dependencies between the input and superficial layers. Using the partial information decomposition framework[13], we estimated modulation of shared dependencies and found that they are reduced within laminar populations, specifically the putative excitatory subpopulations. Surprisingly, we found a strengthening of unique dependencies within the laminar populations, a finding not previously predicted. Crucially, these modulation patterns were also observed across behavioral outcomes (hit vs. miss) that are thought to be mediated by endogenous state fluctuations[14-16]. By "decomposing" the modulation of dependency components and in combination with prior theoretical work[7], our results suggest the following computational model of optimal sensory states that are attained by either task demands or endogenous fluctuations in brain state: enhanced information flow *between* and improved information capacity *within* encoding stages.

42 **MAIN TEXT**

43 Visual cortex has a laminar organization and both sensory computations and information flow
44 patterns are layer specific, forming the building blocks of the ventral visual hierarchy[17-21]. We
45 hypothesized that the deployment of spatial attention strengthens unique inter-layer information
46 transfer between and weakens shared information within the input and superficial layers, both of
47 which are crucial nodes of information flow along the ventral visual hierarchy. To test this, our
48 primary goal was to quantify unique statistical spiking dependencies between the populations of
49 each layer, which requires characterizing the joint spiking activity of the laminar ensemble.
50 Cortical ensembles are highly interconnected, and this can be a source of both shared and unique
51 dependencies between any pair of ensemble members (Fig 1a, S1a-b). Pairwise measures (such
52 as correlations) reflect both unique and shared sources of dependences (Fig 1b,c), whereas
53 network-based methods primarily extract non-shared sources of dependencies (Fig 1d). Using
54 the Partial Information Decomposition[13] framework, we proceeded to use a combination of
55 network-based probabilistic graphical models (PGM)[22] and pairwise regression models to analyze
56 the neurophysiological data and "decompose" the modulation of these dependencies by spatial
57 attention (Fig 1e, see SI Methods). Fitting our data with Dynamic Bayesian Networks (DBN)[23-28],
58 a class of PGMs, allowed unbiased discovery of multi-timescale directed and cyclical
59 dependencies that are Granger causal[29], without making assumptions about their specific nature
60 (linear/nonlinear), direction, or latency (Fig 2a, Fig S2). Directed dependencies are expected to
61 reflect the unique information flow in our data. We estimated the probability of discovered
62 dependencies (referred henceforth as weights of a graph edge) and used it as a measure of their
63 strength[30] (Fig 2a,b). We used a time-shuffled estimate of the edge weights to determine the
64 statistical significance of the discovered dependencies (Fig 2a). We refer to this approach as
65 multi-timescale weighted Dynamic Bayesian Networks (MTwDBN). When tested on a ground truth

66 synthetic network with recurrent connections, MTwDBN approach robustly recovered the
67 dependency structure (Fig 2c-e), and performed significantly better than existing methods,
68 especially when the populations were sparsely sampled (Fig 2f), as is the case in neural
69 recordings. Since the dependencies are discovered without assumption of independence
70 between predictors, edge weights in MTwDBN can be interpreted as an accurate depiction of the
71 structure *and* strength of unique dependencies in a highly interconnected network (Fig 2g).
72
73 We applied the MTwDBN analysis to laminar recordings in visual area V4 (an area strongly
74 modulated by attention[8,11,31-33]) of non-human primates performing an attention demanding
75 orientation change detection task (Fig 3a,b; see Methods and Fig S3). We first analyzed the
76 ensemble of layer-wise populations (Fig 3c) and quantified the net attentional modulation of
77 dependencies across different timescales (Fig 3d). At longer timescales (> 60 ms lag)[33], while
78 attention weakened pairwise dependencies, in agreement with previous findings[34,35], we found a
79 strengthening of unique dependencies. This allowed us to infer (Fig 1e) that it is the shared
80 component of dependencies that is weakened by attention (Fig 3d), thereby providing direct
81 evidence for a previously hypothesized mechanism of perceptual improvement by attention[34,35].
82 Directed dependencies between layers, specifically those between the input and superficial
83 layers, an important link in feedforward processing, were strengthened by attention (Fig 3e). On
84 the other hand, shared dependencies within the input and superficial layers were weakened by
85 attention (Fig 3f). Surprisingly, unique dependencies within the layers were strengthened by
86 attention at most timescales (Fig 3g). Taken together, these results provide the first direct
87 evidence that attention improves unique dependencies both within and across stages of the
88 ventral visual hierarchy, while weakening shared dependencies within these stages (Fig 3h).
89

Neuronal cell-classes in the cortex contribute differentially to information processing[31,32,36]. To test if the above model holds when we allow discovery of cell-class specific dependencies, we next analyzed an ensemble of broad- and narrow-spiking layer-wise populations (Fig 3i) and quantified the net attentional modulation of dependencies across different timescales (Fig 3j). The pattern of net modulation of unique and shared dependencies in this ensemble largely mirrored that discovered in the layer-wise aggregated ensemble, specifically at the longer timescales. Same was the case for modulation of unique dependencies between layers and shared dependencies within layers (consistent across both animals) (Fig 3k,l). Theoretical predictions[7], including our hypotheses, regarding effects of shared fluctuations on information representation apply primarily to excitatory neurons. When we quantified the attention modulation of shared dependencies within layers in a cell-class specific manner, we found that the broad-spiking population (putative excitatory neurons) showed a robust weakening of shared dependencies within layers (consistent across animals) (Fig 3l). On the other hand, the narrow-spiking population (putative inhibitory neurons) showed a distinct pattern, one that was dominated by strengthening of shared dependencies within layers. Consistent with the findings in the aggregated ensemble, unique dependencies within the layers were strengthened by attention at most timescales (Fig 3m). Taken together, these results provide the first direct evidence that attention specifically weakens shared dependencies in the projection populations within encoding stages, in addition to robustly improving unique dependencies both within and across stages of the ventral visual hierarchy (Fig 3n).

To test if the pattern of inter- and intra-layer dependency modulation that we observed is a signature of brain states that are optimal for perceptual behavior, we analyzed the laminar ensemble activity for a subset of trials within the attend-in condition in which the animal was equally likely to correctly detect (Hit) or fail to detect (Miss) the orientation change (Fig 4a,b). Controlling for task and stimulus conditions, these behavioral fluctuations are thought to arise

from endogenous fluctuations such as attentive sampling and arousal changes[14-16,37,38]. The pattern of net modulation of unique and shared dependencies across behavioral outcome (Fig 4c) largely mirrored that discovered across attentive states (Fig 3d) at the longer timescales. Same was the case for modulation of unique dependencies between layers and dependencies within layers (Fig 4d), suggesting that the pattern of enhanced inter-layer unique dependencies and weakened intra-layer shared dependencies is a hallmark of brain states that are optimal for perceptual performance (Fig 4e). Combined with prior theoretical work[7], our results suggest a model of enhanced inter-layer communication and improved intra-layer information capacity during optimal states that are either imposed by task demands or are attained via endogenous fluctuations in brain state.

Our approach demonstrates that the information flow structure in a neural ensemble is robustly described by Dynamic Bayesian Network modeling that is adapted to include multiple lags spanning timescales of interest. This unbiased approach allows us to dive deeper into the causal functional interactions (beyond correlations) in a multivariate system with unknown nonlinear dependencies. Further, the results demonstrate that weighing dependencies using confidence measures provides a more accurate information flow structure that can be utilized to investigate how this structure is modulated by changes in brain and behavioral states. Finally, a novel combination of this approach with pairwise models allows decomposition of changes in unique and shared components of dependencies. Since directed unique dependencies suggest causality and shared ones suggest common inputs, this allows us to gain insights into underlying mechanisms. In contrast, prior correlational studies have quantified net dependency modulation, and thus the underlying mechanisms remained hypothetical [2,9,34].

While approaches such as GLM have been very useful in providing improved phenomenological models of individual neural responses to sensory stimulation in early sensory circuits[39], they are

not suitable for dependency structure learning in a highly recurrent cortical ensemble. Their inherent assumption of the independence of predictor variables can discover spurious dependencies due to shared inputs, and result in a dense and likely inaccurate structure. Since structure learning in DBNs involves determining conditional independence by solving an optimization problem that penalizes density, our approach is ideal for generating a sparse and hence interpretable unique dependency structure in multivariate data.

This class of DBN models does not identify functional consequences of dependencies between populations, such as enhancement or suppression of target activity. Future work in this direction will further enhance the functional interpretability of discovered dependencies. Nonetheless, these models are effective in elucidating structure and modulation of information flow in a multivariate system such as the laminar cortical network.

In general, discovery of dependency structure using our approach is sensitive to the inclusion of relevant neural variables. It is indeed possible that a gross classification of subpopulations as broad and narrow in our study fuses distinct but relevant neural variables. On the other hand, our three and six population laminar DBNs yield consistent patterns regarding modulation of dependencies, suggesting an inclusion of relevant variables for the purposes of this study. It is important to note that while statistical dependencies discovered using our approach imply directed functional connectivity, these do not necessarily imply direct anatomical connectivity: an edge in the structure could reflect indirect anatomical connectivity.

MTwDBN based models provide a quantitative description of information flow pattern in an ensemble. In this study, they provide the first structural description of the attention modulation of dependencies in cortical space and time in a compartmentalized network, such as the laminar

cortical network. They allow us to quantify the unique contributions of activity history and network interactions to information processing in such a neural ensemble. We expect this framework to extend to neuronal ensembles in other parts of the nervous system, and to play an important role in revealing flexible information processing principles in the brain.

## METHODS SUMMARY

**Synthetic Data** Dataset for Figure 1 was constructed by simulating a probabilistic spiking network with causal and shared connectivity among 8 neural variables**.** Dataset for evaluating the MTwDBN pipeline (Fig 2) was generated by simulating a network of stochastically spiking excitatory and inhibitory neurons.

**Electrophysiological Data** 16-channel laminar probes were used to record data from visual area V4 of 2 macaque monkeys as they performed an attention demanding task over 30 sessions total [31]. The probe recorded single units with overlapping receptive fields across the depth of the cortex. Single units were classified using peak-to-trough duration.

**Behavior** Monkeys were trained to perform an attention-demanding change detection task [31].

**Fitting** DBN graph structure was fitted to 200 ms of visual stimulus flashes. Data was prepared using 15 ms bins to fit a DBN with either 21 (3 laminar populations, 7 lags) or 42 nodes (3 lamina x 2 waveform types, 7 lags). Bootstrap methods were used to generate edge weights or confidence measures of edges discovered from the data.

## FIGURE CAPTIONS

**Figure 1 Pairwise- and network-model parameters as a function of unique vs. shared dependencies**. **a** A synthetic ensemble of eight neural variables with two kinds of dependencies – unique or shared – between seven source variables (black) and one target variable (cyan). All interactions are excitatory. Strength of dependencies is determined by model parameters $P_{unique}$ and $P_{shared}$ (see SI Methods). **b** Normalized total mutual information, measured by uncertainty coefficient, as a function of the sum of model parameters ($P_{unique}, P_{shared}$ ) that varied unique and shared components of mutual information in a monotonic way (Fig S1c-d). **c** Coefficients of pairwise model (univariate logistic regression) as a function of $P_{unique}$ and $P_{shared}$. White arrow provides a visual guide for direction of highest change in coefficients. **d** Coefficients of network model (multivariate LASSO regression) as a function of $P_{unique}$ and $P_{shared}$. White arrow provides a visual guide for direction of highest change in coefficients. **e** Schema for utilizing pairwise and network methods for the estimation of total (black) and unique (green) information modulation respectively, and to infer the modulation of shared information (purple). Shaded blocks indicate indeterminate modulation direction of shared information in the network.

**Figure 2 DBN based estimation of multi-timescale weighted unique dependency. a** Analysis flow for multi-timescale weighted DBN (MTwDBN) model fitting. **b** Edge weight of MTwDBN as a function of connection weight in a 2-population simulated network using the pipeline in **a**. **c** Spiking activity of 6 subpopulations in a simulated laminar network. Connectivity is visualized in the overlaid schematic. **d** Directed dependencies (edge in the graph) in the simulated network in **c**, estimated using MTwDBN fitting. **e** Summary graph of dependencies across all timescales from **d**. **f** F score (harmonic mean of precession and recall of dependency structure) as a function of % of population observed. F score was estimated for shuffle corrected weighted DAGs

209 (MTwDBN), weighted DAGs with fixed threshold (FT), or unweighted DAGs. **g** DBN decoder
210 accuracy with different sizes of MTwDBN DAGs. Graph edges for the decoder were sampled from
211 the learned structure either in an unbiased fashion or biased with the edge weights.
212
213 **Figure 3 Modulation of dependencies in a V4 laminar network across attention conditions.**
214 **a** Experimental protocol: Paired Gabor stimuli with different contrasts (see Methods); one stimulus
215 was presented inside the receptive fields (RFs) of the recorded neurons and the other at an
216 equally eccentric location across the vertical meridian. Attention was cued either to the neurons'
217 RFs ("IN") or to the location in the contralateral visual hemifield ("AWAY"). Orientation of one of
218 the two stimuli changed at a random time. Monkey was rewarded for detecting the change by
219 making a saccade to the corresponding location. Task difficulty was controlled by the magnitude
220 of orientation change. **b** Recording approach: laminar recordings in visual area V4. **c** Neural
221 populations used for MTwDBN analysis. Current source density analysis identified different layers
222 (*superficial*, *input*, *deep*), and isolated single units were assigned to one of these layers (see SI
223 Methods). **d** *Top*: MTwDBN-based modulation (green) of all unique dependencies between the
224 laminar populations. Modulation of the same dependencies as estimated by logistic regression
225 (black). *Bottom*: PID framework-based estimated modulation sign of shared dependencies (see
226 schema in Fig 1e). Thicker line along the time axis indicates the timescales of attentional
227 modulation in prior studies[34,35]. **e** Sign of average modulation of unique dependencies between
228 layers. **f** Sign of average modulation of shared dependencies within layers. **g** Sign of average
229 modulation of unique dependencies within layers. **h** Summary of dependency modulation pattern.
230 **i** Neural populations used for laminar MTwDBN analysis. Isolated single units were classified as
231 broad and narrow based on peak-to-trough duration in their average spike shape (see SI
232 Methods). **j-n** Same as **d-h**, but for populations shown in **i**. I: input layer; S: superficial layer; *M1*,

233 *M2*: subject-wise dependencies; *broad*, *narrow*: cell-class specific dependencies. See Fig S4 for
234 modulation indices in **e-g**, **k-n**.
235
236 **Figure 4 Modulation of dependencies in a V4 laminar network across behavioral outcomes**
237 **at perceptual threshold. a** Example session showing performance as a function of task difficulty.
238 (orange box: threshold orientation change at which the animal was equally likely to correctly
239 detect (Hit) or fail to detect (Miss) the change). **b** Laminar populations used for multi-lag analysis.
240 **c** Modulation magnitude (top) and sign (bottom) of all unique (green) and total (black) laminar
241 dependencies in **b** across Hits and Misses at perceptual threshold. Estimated modulation sign of
242 shared dependencies (bottom, see Fig 1e). **d** Modulation sign of between (BL) and within (WL)
243 layer dependencies. See Fig S5 for modulation indices. **e** Summary of laminar dependency
244 modulation pattern.
245
246 **Figure S1 Information decomposition in a multi-variate system**
247 **a** Partial information decomposition (PID) framework based definition of types of information that
248 multiple sources can have about a target.
249 **b** Application of pairwise and network-based statistical models for approximate information
250 decomposition in a multivariate system. Panel on the right illustrates interpretation of the
251 modulation of these dependencies using the PID framework.
252 **c, d** Reduction in the proportion of total entropy (or information fraction) as a function of
253 parameters ($P_{unique}, P_{shared}$) that control unique and shared information in the model. For the sake
254 of computational efficiency, information fraction estimation as a function of $P_{shared}$ (d) was
255 performed using a subset of variables in the simulated network.
256

**Figure S2 Additional MTwDBN performance metric for synthetic data**

Recall, precision and F score measures for weighted DAGs in Fig 2d as a function of number of lags.

**Figure S3 Laminar recordings in area V4**

**a** Stacked contour plot showing spatial receptive fields (RFs) along the laminar probe from an example session. The RFs are well aligned, indicating perpendicular penetration down a cortical column. Zero depth represents the center of the input layer as estimated with current source density (CSD) analysis. **b** CSD displayed as a colored map. The x-axis represents time from stimulus onset; the y-axis represents cortical depth. The CSD map has been spatially smoothed for visualization. **c** An example trial showing single-unit activity across the cortical depth in the attend-in condition. The time axis is referenced to the appearance of the fixation spot. Spikes (vertical ticks) in each recording channel come from the single unit with the highest spike rate in this trial. The gray boxes depict stimulus presentation epochs. In this particular trial, 8 sample stimuli with different contrasts were presented, followed by a target stimulus flash with an orientation change that the monkey responded to correctly. Spike waveforms for two example narrow (cyan) and broad (orange) single units are shown.

**Figure S4 Modulation of dependencies in a V4 laminar network across attention conditions.** MTwDBN-based modulation (green) of all unique dependencies between the laminar populations. Modulation of the same dependencies as estimated by logistic regression (black). Open circles indicate no significant edges for a given time lag. Schematic at the top of every column depicts the populations used for identifying laminar dependency structure within (left) and between (right) cortical layers. *broad, narrow*: cell-class specific populations identified by spike shape (see SI Methods).

**Figure S5 Modulation of dependencies in a V4 laminar network across behavioral outcomes at perceptual threshold**. MTwDBN-based modulation (green) of all unique dependencies between the laminar populations. Modulation of the same dependencies as estimated by logistic regression (black). Schematic at the top of every column depicts the populations used for identifying laminar dependency structure within (left) and between (right) cortical layers.

## AUTHOR CONTRIBUTIONS

AD and MPJ conceptualized the project. ASN collected the electrophysiological data. AD and MPJ generated synthetic data. AGS and AD analyzed the data. MPJ supervised the project. AGS, AD, ASN and MPJ wrote the manuscript.

## ACKNOWLEDGEMENTS

This research was supported by NIH R00 EY025026, NIH R21 MH126072 and SFARI 875855 to MPJ, NARSAD Young Investigator Grant, Ziegler Foundation Grant, Yale Orthwein Scholar Funds, NIH R21 MH126072 and SFARI 875855 to ASN, NSF GRFP fellowship to AGS and by NEI core grant for vision research P30 EY026878 to Yale University. We thank Steve Chang and Weikang Shi for helpful comments on the manuscript.

## REFERENCES

1 Petersen, S. E. & Posner, M. I. The attention system of the human brain: 20 years after. *Annual review of neuroscience* **35**, 73-89, doi:10.1146/annurev-neuro-062111-150525 (2012).

2 Ruff, D. A. & Cohen, M. R. Attention Increases Spike Count Correlations between Visual Cortical Areas. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **36**, 7523-7534, doi:10.1523/JNEUROSCI.0610-16.2016 (2016).

3 Briggs, F., Mangun, G. R. & Usrey, W. M. Attention enhances synaptic efficacy and the signal-to-noise ratio in neural circuits. *Nature* **499**, 476-480, doi:10.1038/nature12276 (2013).

4 Womelsdorf, T. & Fries, P. The role of neuronal synchronization in selective attention. *Current opinion in neurobiology* **17**, 154-160, doi:10.1016/j.conb.2007.02.002 (2007).

5 Bosman, C. A. *et al.* Attentional stimulus selection through selective synchronization between monkey visual areas. *Neuron* **75**, 875-888, doi:10.1016/j.neuron.2012.06.037 (2012).

6 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, doi:10.1126/science.1171402 (2009).

7 Zohary, E., Shadlen, M. N. & Newsome, W. T. Correlated neuronal discharge rate and its implications for psychophysical performance. *Nature* **370**, 140-143, doi:10.1038/370140a0 (1994).

8 Mitchell, J. F., Sundberg, K. A. & Reynolds, J. H. Spatial attention decorrelates intrinsic activity fluctuations in macaque area V4. *Neuron* **63**, 879-888, doi:10.1016/j.neuron.2009.09.013 (2009).

9 Cohen, M. R. & Maunsell, J. H. R. Attention improves performance primarily by reducing interneuronal correlations. *Nat Neurosci* **12**, 1594-1600, doi:10.1038/nn.2439 (2009).

10 Desimone, R. & Duncan, J. Neural mechanisms of selective visual attention. *Annual review of neuroscience* **18**, 193-222 (1995).

11 McAdams, C. & Maunsell, J. Effects of Attention on Orientation-Tuning Functions of Single Neurons in Macaque Cortical Area V4. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **19**, 431-441 (1999).

12 Reynolds, J., Pasternak, T. & Desimone, R. Attention increases sensitivity of V4 neurons. *Neuron* **26**, 703-714 (2000).

13 Kohn, A., Coen-Cagli, R., Kanitscheider, I. & Pouget, A. in *Annual review of neuroscience* Vol. 39 237-256 ( Annual Reviews 2016).

14 Fiebelkorn, I. C., Pinsk, M. A. & Kastner, S. in *Neuron* Vol. 99 842-853.e848 (2018).

15 Helfrich, R. F. *et al.* in *Neuron* Vol. 99 854-865.e855 (2018).

16 Fiebelkorn, I. C., Saalmann, Y. B. & Kastner, S. in *Current Biology* Vol. 23 2553-2558 (Elsevier, 2013).

17 Douglas, R. J. & Martin, K. A. C. Neuronal circuits of the neocortex. *Annual review of neuroscience* **27**, 419-451, doi:10.1146/annurev.neuro.27.070203.144152 (2004).

18 Douglas, R. J. & Martin, K. A. C. Recurrent neuronal circuits in the neocortex. *Current biology : CB* **17**, R496-500, doi:10.1016/j.cub.2007.04.024 (2007).

19 D'Souza, R. D. & Burkhalter, A. in *Front Neuroanat* Vol. 11 71 (2017).

20 Gattass, R., Galkin, T. W., Desimone, R. & Ungerleider, L. G. in *The Journal of comparative neurology* Vol. 522 1941-1965 (John Wiley & Sons, Ltd, 2014).

21 Ungerleider, L. G., Galkin, T. W., Desimone, R. & Gattass, R. Cortical connections of area V4 in the macaque. *Cerebral cortex (New York, NY : 1991)* **18**, 477-499, doi:10.1093/cercor/bhm061 (2008).

22 Koller, D. & Friedman, N. *Probabilistic graphical models: principles and techniques*. (MIT press, 2009).

23 Murphy, K. P. & Russell, S. Dynamic bayesian networks: representation, inference and learning. (2002).
24 Smith, V. A., Jarvis, E. D. & Hartemink, A. J. Evaluating functional network inference using simulations of complex biological systems. *Bioinformatics* **18**, S216-S224 (2002).
25 Zhang, L., Samaras, D., Alia-Klein, N., Volkow, N. & Goldstein, R. in *Advances in neural information processing systems.* 1593-1600.
26 Rajapakse, J. C. & Zhou, J. Learning effective brain connectivity with dynamic Bayesian networks. *Neuroimage* **37**, 749-760 (2007).
27 Smith, V. A., Yu, J., Smulders, T. V., Hartemink, A. J. & Jarvis, E. D. Computational inference of neural information flow networks. *PLoS computational biology* **2** (2006).
28 Eldawlatly, S., Zhou, Y., Jin, R. & Oweiss, K. G. On the use of dynamic Bayesian networks in reconstructing functional neuronal networks from spike train ensembles. *Neural Comput* **22**, 158-189, doi:10.1162/neco.2009.11-08-900 (2010).
29 Seth, A. K., Barrett, A. B. & Barnett, L. Granger causality analysis in neuroscience and neuroimaging. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **35**, 3293-3297, doi:10.1523/JNEUROSCI.4399-14.2015 (2015).
30 Friedman, N., Goldszmidt, M. & Wyner, A. Data analysis with Bayesian networks: A bootstrap approach. *arXiv preprint arXiv:1301.6695* (2013).
31 Nandy, A. S., Nassi, J. J. & Reynolds, J. H. Laminar Organization of Attentional Modulation in Macaque Visual Area V4. *Neuron* **93**, 235-246, doi:10.1016/j.neuron.2016.11.029 (2017).
32 Mitchell, J. F., Sundberg, K. A. & Reynolds, J. H. Differential attention-dependent response modulation across cell classes in macaque visual area V4. *Neuron* **55**, 131-141, doi:10.1016/j.neuron.2007.06.018 (2007).
33 Cohen, M. R. & Maunsell, J. H. R. Using neuronal populations to study the mechanisms underlying spatial and feature attention. *Neuron* **70**, 1192-1204, doi:10.1016/j.neuron.2011.04.029 (2011).
34 Mitchell, J. F., Sundberg, K. A. & Reynolds, J. H. in *Neuron* Vol. 63 879-888 (2009).
35 Cohen, M. R. & Maunsell, J. H. R. in *Nat. Neurosci.* Vol. 12 1594-1600 (2009).
36 Pettine, W. W., Steinmetz, N. A. & Moore, T. Laminar segregation of sensory coding and behavioral readout in macaque V4. *Proceedings of the National Academy of Sciences of the United States of America* **116**, 14749-14754, doi:10.1073/pnas.1819398116 (2019).
37 McGinley, M. J., David, S. V. & McCormick, D. A. in *Neuron* Vol. 87 179-192 (2015).
38 Landau, A. N. & Fries, P. Attention samples stimuli rhythmically. *Curr Biol* **22**, 1000-1004, doi:10.1016/j.cub.2012.03.054 (2012).
39 Pillow, J. W. *et al.* Spatio-temporal correlations and visual signalling in a complete neuronal population. *Nature* **454**, 995-999, doi:10.1038/nature07140 (2008).

## Extended Methods

*Partial Information Decomposition*

Information theory does not provide a complete description of the informational relationships between variables in a system composed of three or more variables. The information $\mathrm{I(T; S1, S2)}$ that two 'source' variables $\mathrm{S1}$ and $\mathrm{S2}$ hold about a third 'target' variable $\mathrm{T}$ decomposes into four parts: (i)$\mathrm{U(T; S1|S2)}$, the unique information that only $\mathrm{S1}$ (out of $\mathrm{S1}$ and $\mathrm{S2}$) holds about $\mathrm{T}$; (ii) $\mathrm{U(T; S2|S1)}$, the unique information that only $\mathrm{S2}$ holds about $\mathrm{T}$; (iii) $\mathrm{R(T; S1, S2)}$, the redundant information that both $\mathrm{S1}$ and $\mathrm{S2}$ hold about $\mathrm{T}$; and (iv) $\mathrm{S(T; S1, S2)}$, the synergistic information about $\mathrm{T}$ that only arises from knowing both $\mathrm{S1}$ and $\mathrm{S2}$ (see Fig S1a). The set of quantities $\{\mathrm{U(T; S1\,|S2), U(T;\ S2|S1), R(T; S1, S2), S(T; S1, S2)}\}$ is called a 'partial information decomposition' related as follows:

$$\mathrm{I(T; S1, S2) = U(T; S1|S2) + U(T;\ S2|S1\,) + S(T; S1, S2) + R(T; S1, S2)}$$

$$\mathrm{I(T; S1\,) = U(T; S1|S2) + R(T; S1\,, S2)}$$

$$\mathrm{I(T; S2) = U(T; S2|S1\,) + R(T; S1\,, S2)}$$

Thus, the Partial Information Decomposition (PID)[1] framework characterizes the mutual information between variables by decomposing it into unique, shared, and synergistic components.

In a multi-neuronal network, unique (and synergistic) components of the mutual information between neural variables due to causal interactions (unique and synergistic) are captured by directed statistical dependencies. On the other hand, shared components of mutual information between neural variables correspond to shared neuronal inputs, and the sign of their modulation can be estimated from the modulation of unique and total mutual information (Fig S1b).

*Information Decomposition with Pairwise and Network Models*

a. Generation of synthetic data

A synthetic network was generated in which unique and shared dependencies between variables were controlled. Eight variables were randomly initialized to values of 0 or 1. One variable was assigned as the target, the other seven as sources. One source was designated as the unique source. Samples were generated in the following manner: 1) the value of all variables was set to 1 with probability $P_{shared}$ and to 0 with probability $1 - P_{shared}$ and 2) the values of the target and unique source were set to 1 with probability $P_{unique}$ and to 0 with probability $1 - P_{unique}$. This created a shared dependency between all eight variables and a single unique dependency between the target and the unique source. Two thousand (2000) samples each were generated with $P_{shared}$, $P_{unique}$ ∈ {0.1, 0.2, 0.3, 0.4}. The total normalized mutual information between the unique source and the target variable was calculated by the uncertainty coefficient:

$$U(target|source) = \frac{H(target) - H(target|source)}{H(target)}$$

, where $H$ is entropy. Total entropy of target and unique source variables, $H(target, unique\ source)$, was additionally calculated for $P_{unique}$ ∈ {0.1, 0.2, 0.3, 0.4} and $P_{shared}$ ∈ {0.1, 0.4}. Values were normalized to the entropy for ($P_{unique}$ , $P_{shared}$) = (0.1, 0.1) to represent the change in total entropy (information fraction). Similarly, total entropy was calculated for 100 random subsets of three source variables plus the target variable for $P_{unique}$ ∈ {0.1, 0.4} and $P_{shared}$ ∈ {0.1, 0.2, 0.3, 0.4} and normalized to entropy for ($P_{unique}$ , $P_{shared}$) = {0.1, 0.1}.

b. Pairwise models

Univariate logistic regression was performed between the unique source and target variable using the statsmodel python package[1] for all values of $P_{shared}$ and $P_{unique}$.

c. Network models

Multivariate regression was performed between all seven source variables and the target variable using the statsmodel python package using L1 penalty[2-4].

*Multi-timelag Weighted Dynamic Bayesian Network (MTwDBN) Analysis Pipeline*

*a. Fitting Dynamic Bayesian Network models*

A Dynamic Bayesian Network (DBN) framework was used to learn dependencies between neural populations[5]. The pgmpy python package[6] (https://github.com/pgmpy/pgmpy) and custom-written python code was used to fit all DBN models. Each binned-and-sliced data table was first bootstrapped *B* times. For each session/condition, a hill climb tabu-search with a history window of 7 was performed 120 times, each from a unique random starting graph, to find a suitable fitting directed acyclic graph (DAG)[7]. The Akaike information criterion (AIC) was used[8,9] as the scoring metric in the tabu-search. The variables associated with the latest time slice (0 ms) were termed the *effect* variables; all others were termed potential *cause* variables. The search was restricted to DAGs where edges can only be incident on an *effect* variable. Edges between *effect* variables were allowed as they may capture dependencies at timescales shorter than our chosen bin widths, however such edges have no causal interpretation and were excluded from further analysis. The resulting DAGs were termed *unweighted*, as their edges (dependencies) are described in a present/absent manner. Of the 120 starting points, only the DAG with highest AIC score was used for further analysis. This resulted in *B unweighted* DAGs per session/condition.

463 *b. Estimation of weighted DAG*

464 The *B* DAGs from each session/condition were used to estimate weighted DAGs[10]. A pool of 100
465 weighted DAGs were estimated by taking 100 bootstrap samples from the *B unweighted* DAGs
466 and averaging: the weight $\in (0,1)$ for each edge corresponds to the proportion of the *B unweighted*
467 DAGs where the edge was present. This resulted in 100 weighted DAGs for each
468 session/condition.

469

470 *c. Testing significance of DAG edges*

471 To test for significance of the discovered edges, the 100 weighted DAGs were compared to 100
472 control DAGs that were generated in the same manner as above but with time-shuffled data. The
473 binned-and-sliced data tables were shuffled as follows: for each row, permute the data from each
474 of the 7 time slices which arose from the same population. Finally, permute each column
475 (corresponding to each population/time slice combination). To test for significance, a distribution
476 of weights was generated by combining weights for a given edge across all sessions (conditions
477 treated separately). The distribution of edge weights from unshuffled data were compared to the
478 distribution of edge weights from the shuffled data using a one-sided Mann Whitney U Test. If the
479 distribution from unshuffled data had significantly higher values than from shuffled data ($p<0.05$),
480 the edge was marked as significant for that condition and said to have survived time shuffling.

481

482 *MTwDBN Validation*

483 *a. Synthetic Neural Network Models*

484 Synthetic neural network models were constructed using stochastic spiking neurons[11,12].
485 Individual neurons in the model were treated as coupled, continuous-time, two-state (active and
486 quiescent) Markov processes. The active state represents a neuron firing an action potential and
487 its accompanying refractory period, whereas the quiescent states represent a neuron at rest. The

transition probability for the i-th neuron to decay from active to quiescent state in time *dt* was $P_i(\text{active} \rightarrow \text{quiescent}) = \alpha_i\, \partial(dt)$, where $\alpha_i$ represented the decay rate of the active state of the neuron. Parameter $\alpha_i$ sets the upper bound on firing rate of the stochastically spiking neuron, akin to a refractory period. The transition probability for the i-th neuron to change from quiescent to active state (i.e., spike) was $P_i(\text{quiescent} \rightarrow \text{active}) = \beta_i G(S_i)\partial(dt)$ [11]. This caused the firing probability to be a function of the input, with $\beta_i$ as its peak value. Parameter $S_i$ was the total synaptic input to neuron i, given as $S_i(t) = N_i(t) + I_i(t)$, where $N_i$ was the net input from other neurons in the local network and $I_i$ was the net external input to the neuron. The network input was $N_i(t) = \sum_j w_{ij}\, A_j(t)$, where $w_{ij}$ are the weights of the synapses. The activity variable $A_j(t)$ was set to one if the jth neuron was active at time *t* and zero otherwise. The model neurons had no intrinsic capacity to oscillate because the inter-spike interval was the sum of two independent exponential random variables with parameters $\alpha_i$ and $\beta_i G(S_i)$, respectively.

The model parameters were chosen as follows: Excitatory (E) and inhibitory (I) neurons in the network were differentiated based on two model parameters: $\alpha_E = 0.075$ ms; $\alpha_I = 0.4$; ms and $\beta_E = 1$; $\beta_I = 2$.

*1. Two-Population Model*

The model consisted of two excitatory neurons (each representing a distinct population). There was a single synaptic connection from the first neuron to the second neuron. Ten such two-population models were constructed with synaptic weights ∈ (0.5, 1.0, …, 9.0, 9.5). 1000 trials of spiking data each 100 ms long were simulated for each model.

### 2. *Six-population Laminar Model*

Laminar models consisted of 45 neurons in total, 15 in each cortical layer (superficial, input, deep). Each layer contained 10 excitatory and 5 inhibitory neurons, giving a total of 6 populations (3 layers x 2 neuron types). The network topology for synaptic connectivity is depicted in Fig 2c. The weights $w_{ij}$ for synaptic connections are 1 for the inter-laminar connections, 1.5 for intra-laminar connections where the presynaptic unit is an E unit, and -2 for intra-laminar connections where the presynaptic unit is an I unit. 1000 trials of spiking data, each 2000 ms long, were simulated using this model.

## b. *Preprocessing for MTwDBN analysis*

Data from single units was grouped by population in each model simulation. The "multi-unit" spiking activity of these populations was used for the analysis. Before aggregating activities of single units into populations, *d* % (*d* being a pre-selected number, see *Effect of sub-sampling in synthetic laminar model*) of single units in each population were dropped (not used for analysis) from the laminar model data. The data from each trial were discretized into 1.2ms bins of either 0, 1, or 2 to denote no spikes, one spike, or multiple spikes in a time bin. The data were lagged 2 times to give 3 time slices (2.4, 1.2, 0 ms). The data from all trials were concatenated together to generate a single data table with 6 or 18 columns (2 or 6 populations x 3 time slices). The binned and discretized spiking activity of a single population in a single time slice was viewed as a variable in either the pairwise regression or DBN framework. Data with structure as produced by this preprocessing step are termed binned-and-sliced data tables.

535 *c. Analysis: Relation between synaptic weights and DAG edge weights in two-population*
536 *models*

537 Binned-and-sliced data tables from each two-population model were passed separately through
538 the MTwDBN pipeline ($B$ = 200) to generate 100 weighted DAGs each. Weights were obtained
539 by averaging across the 200 *unweighted* DAGs. The 100 weighted DAGs were used to compute
540 95% confidence intervals for these weights.

541

542 *d. Analysis: Effect of sub-sampling on recovering ground-truth in synthetic laminar model*

543 To assess the effect of sub-sampling neural data on MTwDBN outputs, *d* % (d ∈ {0, 20, 40, 60,
544 80}) of neurons in each of 6 populations were dropped from the laminar model before
545 preprocessing (*Preprocessing Synthetic Neural Network Data)* resulting in five binned-and-sliced
546 data tables. Each data table was passed through the MTwDBN pipeline ($B$ = 50), generating 50
547 *unweighted* DAGs and 100 weighted DAGs. Edges were considered significant either by surviving
548 time shuffling (MTwDBN) or by passing a fixed threshold of 0.5 (weighted + FT). To measure how
549 well they conformed to ground truth connectivity, an F-score was calculated for models
550 (MTwDBN, weighted + FT, unweighted) fitted to a given subsampled dataset. We use a
551 framework[5] where edges are considered regardless of the time lag in which they appear. The F-
552 Score is defined as: $F = \frac{RP}{R+P}$ , with $R = \frac{C}{C+M}$, $P = \frac{C}{C+I}$, and $C$ = # connections inferred by DBN
553 which are in the ground truth, $M$ = # connections in ground truth not inferred by DBN, $I$ =#
554 connections inferred by DBN which are not in the ground truth. R and P refer to Recall and
555 Precision.

556

557

558 *e. Analysis: Effect of number of time lags on recovering ground-truth in synthetic laminar*
559 *model*

560 To assess the impact of including different number of time lags on recovering ground-truth
561 dependencies, the synthetic laminar data was preprocessed ($d = 0$) with # of lags $\in \{3, 4, 5, 6, 7\}$.
562 Each binned-and-sliced data table was run independently through the MTwDBN pipeline ($B = 50$)
563 and F-scores, Recalls, and Precisions calculated.

564

565 *f. Analysis: Validating Predictive Power of Edge Weights in Synthetic Laminar Model*

566 To validate whether the edge weights contain additional information about the population
567 dependencies above unweighted edge weights, the binned-and-sliced data table from the
568 synthetic laminar network ($d = 0$, no neurons dropped) was divided into two data tables row-wise,
569 $data_{train}$ (95% of data) and $data_{test}$ (5% of data). $data_{train}$ was bootstrapped 10 times and separately
570 passed through the MTwDBN pipeline ($B = 50$) to generate 10 x 100 weighted DAGs. The set of
571 edges in the weighted graph that survived time-shuffling are denoted $E_{set}$. Subsampled
572 *unweighted* DAGs containing $n$ edges (6 or 13) were obtained by sampling without replacement
573 from $E_{set}$ by either sampling uniformly (UW, unweighted) or sampling using the DAG edge weights
574 as sampling weights (W, weighted). Models UW and W were used for prediction by fitting
575 parameters to 12,000 samples from $data_{test}$ by maximum likelihood estimation[13]. From the
576 remaining samples in $data_{test}$ not used for parameter fitting, 4,000 samples were used for
577 validation. For each sample in the validation set, one of the *effect* variables was chosen at random
578 and both UW and W were used to predict the value of this *effect* variable from the *cause* variables.
579 The prediction accuracies of the two models were compared to the true effect variable value using
580 the M score[14] . This was repeated for 100 variations of UW and W (created by varying the random
581 seed used to sample edges), and the resulting 100 M scores were compared using a two-tailed

582 paired t-test for both $n$=6 and $n$=13. The W model is considered to outperform UW in prediction if
583 $M_W > M_{UW}$ and $p < 0.05$.
584
585 *Attention Data*
586 *a.* Behavioral task
587 Well-isolated single units were recorded from area V4 of two rhesus macaques during an
588 attention-demanding orientation change detection task. The task design and the experimental
589 procedures are described in detail in a previous study[15,16]. While the monkey maintained fixation,
590 two oriented Gabor stimuli were flashed on for 200 ms and off for variable intervals (randomly
591 chosen between 200 and 400 ms). The contrast of the stimulus was randomly chosen from a
592 uniform distribution of 6 contrasts (c = [10%, 18%, 26%, 34%, 42%, and 50%]). One of the stimuli
593 was located at the receptive field overlap region (*Attend In*) and the other at an equally eccentric
594 location across the vertical meridian (*Attend Away*). At the beginning of a block of trials, we
595 presented instruction trials where the monkey was spatially cued to the covertly attend to one of
596 two stimulus locations. One of the two stimuli changed in orientation at an unpredictable time
597 (minimum 1s, maximum 5s, mean 3s). The monkey was rewarded for making a saccade to the
598 location of orientation change. 95% of the orientation changes occur at the cued location, and 5%
599 occur at the uncued location (foil trials). We observed impaired performance and slower reaction
600 times for the foil trials, suggesting that the monkey was indeed using the spatial cue to perform
601 the task. The difficulty of the task was controlled by changing the degree of orientation change
602 (randomly chosen from the following: 1°, 2°, 3°, 4°, 6°, 8°, 10°, and 12°). If no change occurred
603 before 5 s, the monkey was rewarded for holding fixation (catch trial, 13% of trials).
604
605
606

607 *b.* Electrophysiological recording

608 While the monkey was performing the attention task, we used artificial dura chambers to facilitate
609 the insertion of 16-channel linear array electrodes ("laminar probes", Plexon, Plexon V-probe) into
610 cortical sites near the center of the prelunate gyrus. Neuronal signals were recorded, filtered, and
611 stored using the Multichannel Acquisition Processor system (Plexon). Neuronal signals were
612 classified as either isolated single units or multiunit clusters by the Plexon Offline Sorter program.
613 For the data collected from linear array electrodes, we used current source density analysis to
614 identify the superficial (Layers 1-3), input (Layer 4), and deep (Layers 5 and 6) layers of the cortex
615 based on the second derivative of the flash-triggered LFPs[15,17]. Cell bodies of single units with bi-
616 phasic action potential waveforms were assigned to the same layer in which the electrode channel
617 was situated during recordings. Units that had tri-phasic waveforms or other shapes were
618 excluded from analyses. Units with peak-to-trough duration greater than 225$\mu$s were classified as
619 broad-spiking putative excitatory neurons; units with peak-to-trough duration less than 225$\mu$s
620 were classified as narrow-spiking putative inhibitory neurons. Extracellular data were collected
621 over 32 sessions (23 sessions in monkey A, 9 in monkey C) yielding 337 single units in total. Unit
622 yield per session was considerably higher in monkey C than monkey A, resulting in a roughly
623 equal contribution of both monkeys toward the population data.

624

625 *c. Data Selection*

626 All analyses in this study were performed on spiking data during an interval of 60-260 m after
627 stimulus onset excluding orientation changes. Only single units whose spike waveforms were
628 successfully classified as broad or narrow and for whom the layer identity could be successfully
629 discerned were used in the analysis. There were 29 sessions which had one or more such units
630 recorded. For layer-wise analyses, only sessions with at least one unit from each layer was
631 included (18 sessions). For broad- and narrow-spiking layer-wise (layer-class) analyses, only

632 sessions with at least one unit from two or more populations were included (27 sessions). Pairwise
633 and network-based dependency analysis were performed on each session separately.
634
635 *d. Analysis: Attention Conditions*
636 Data from *Attend In* and *Attend Away* trials were analyzed independently. For each attention
637 condition, only data from trials where the animal successfully detected the orientation change or
638 from catch trials where the animal maintained fixation were used.
639
640 *e. Analysis: Behavioral Performance*
641 We fit the behavioral data with a logistic function and defined the threshold condition as the
642 orientation change that was closest to the 50% threshold of the fitted psychometric function for
643 that session. This subset of trials from within the attend-in condition in which the animal was
644 equally likely to correctly detect (*Hit*) or fail to detect (*Miss*) the orientation change was used for
645 our analysis. Data from *Hit* and *Miss* trials were analyzed independently.
646
647 *f. Preprocessing for MTwDBN analysis*
648 Single units were grouped according to neocortical layer (superficial/input/deep) for layer-wise (3
649 populations) analyses and additionally by spike waveform (narrow/broad) for layer+class (3x2
650 populations) analyses. The "multi-unit" spiking activity of these populations was used for the
651 analysis. The data from each stimulus presentation (60 - 260 ms after stimulus onset) were
652 discretized into 15 ms bins and 6 lags to give 7 time slices (-90, -75, -60, -45, -30, -15, 0 ms). The
653 spiking activity of each population in each bin was discretized to 1 or 0 to denote if there were
654 spikes or not. The data from all stimulus presentations in a session/condition combination were
655 concatenated together. Data tables had 21 columns for layer-wise analyses (3 layers x 7 time

slices) and 42 columns for layer-class analyses (6 layer-class populations x 7 time slices). The binned and discretized spiking activity of a single population in a single time slice was viewed as a variable in either the pairwise regression or DBN framework. Data from each session/condition were preprocessed separately. To keep analyses consistent across conditions being compared (*Attend-In vs. Away* OR *Hit vs. Miss*), the size of the bootstraps was equal to the maximum number of rows of the two conditions.

### *g.* Estimation of Condition Modulated Edges

Binned and sliced data tables from each session/condition were independently passed through the MTwDBN pipeline ($B$ = 200) to generate 100 weighted DAGs each. Estimation of condition modulated edges was only performed for those that survived time shuffled in at least one of the conditions to be compared (*Attend In* or *Away*, *Hit* or *Miss*). For each session where both *cause* and *effect* populations were recorded, 5000 pairs of weighted DAGs were bootstrapped from the two conditions to be compared. For each pair, attention modulation indices (AMI) and hit modulation indices (HMI) were calculated as:

$$AMI = \frac{(Edge\ Weight)_{IN} - (Edge\ Weight)_{AWAY}}{(Edge\ Weight)_{IN} + (Edge\ Weight)_{AWAY}}$$

$$HMI = \frac{(Edge\ Weight)_{HIT} - (Edge\ Weight)_{MISS}}{(Edge\ Weight)_{HIT} + (Edge\ Weight)_{MISS}}$$

This resulted in 5000 modulation values for each edge that survived time-shuffling and each recording session where both *cause* and *effect* populations were recorded.

### *h. Pairwise Dependency Analysis*

Total dependency weights between neocortical populations were estimated using univariate logistic regression. This analysis was only performed on edges that survived time-shuffling in the MTwDBN analysis. For each session where both *cause* and *effect* populations were recorded, 5000 samples were bootstrapped from the binned-and-sliced data tables (see *Data Preprocessing*) for each condition separately. To keep analyses consistent across *Attend-In/Away* and *Hit/Miss,* the size of the bootstraps was equal to the maximum number of rows of the two conditions to be compared. Logistic regression was performed separately from each *cause* variable to an *effect* variable using the statsmodel python package, BGFS solver[1]. The absolute value of the $\beta_1$ coefficient was treated as the total dependency weight to mirror the unsigned dependencies discovered using DBNs. This resulted in 5000 pairs of $\beta_1$ across the conditions to be compared. For each pair, attention modulation indices (AMI) and hit modulation indices (HMI) were calculated as:

$$AMI = \frac{|\beta_1^{\text{in}}| - |\beta_1^{\text{away}}|}{|\beta_1^{\text{in}}| + |\beta_1^{\text{away}}|}$$

$$HMI = \frac{|\beta_1^{\text{hit}}| - |\beta_1^{\text{miss}}|}{|\beta_1^{\text{hit}}| + |\beta_1^{\text{miss}}|}$$

This resulted in 5000 modulation values for each edge that survived time-shuffling in the MTwDBN analysis and each recording session where both *cause* and *effect* populations were recorded.

### *i. Calculating Confidence Intervals of Modulation Indices*

Modulation indices (either attention or hit) were grouped according to time lag and additionally whether they were between layers (all layers or input ⇔ superficial only) or within layers (all layers, input only, superficial only). For layer-class analyses, they could be additionally grouped according to broad or narrow waveforms. Due to the large sample sizes generated from bootstrapping, hypothesis testing discovers all modulations to be significantly different from zero ($p = 0$). Therefore, an estimation statistics approach was used to estimate confidence intervals of modulation indices. The mean modulation index and bias-corrected and accelerated bootstrap 95% confidence intervals were calculated in python (scipy.stats.bootstrap, modified to allow for setting size of bootstrap). To estimate confidence intervals more conservatively given the large sample sizes, the number of resamples and the size of each bootstrap were each set to 5000.

## EXTENDED METHODS REFERENCES


1 Seabold, S. & Perktold, J. in *Proceedings of the 9th Python in Science Conference.* 61 (Austin, TX).

2 Tibshirani, R. Regression shrinkage and selection via the lasso: a retrospective. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **73**, 273-282 (2011).

3 Pedregosa, F. *et al.* Scikit-learn: Machine Learning in Python. *J Mach Learn Res* **12**, 2825-2830 (2011).

4 Kuhn, M. & Johnson, K. *Applied predictive modeling*. (Springer, 2013).

5 Eldawlatly, S., Zhou, Y., Jin, R. & Oweiss, K. G. On the use of dynamic Bayesian networks in reconstructing functional neuronal networks from spike train ensembles. *Neural computation* **22**, 158-189 (2010).

6 Ankan, A. a. P., Abinash. in *Proceedings of the 14th Python in Science Conference (SCIPY 2015)* (2015).

7 Koller, D. & Friedman, N. *Probabilistic graphical models: principles and techniques*. (MIT press, 2009).

8 Akaike, H. in *Selected papers of hirotugu akaike* 199-213 (Springer, 1998).

9 Liu, Z., Malone, B. & Yuan, C. in *BMC bioinformatics.* 1-16 (Springer).

10 Friedman, N., Goldszmidt, M. & Wyner, A. Data analysis with Bayesian networks: A bootstrap approach. *arXiv preprint arXiv:1301.6695* (2013).

11 Jadi, M. P. & Sejnowski, T. J. Cortical oscillations arise from contextual interactions that regulate sparse coding. *Proceedings of the National Academy of Sciences* **111**, 6780-6785 (2014).

12 Wallace, E., Benayoun, M., van Drongelen, W. & Cowan, J. D. Emergent oscillations in networks of stochastic spiking neurons. *PLoS One* **6**, e14804, doi:10.1371/journal.pone.0014804 (2011).

13 Neapolitan, R. E. *Learning bayesian networks*. Vol. 38 (Pearson Prentice Hall Upper Saddle River, NJ, 2004).

14 Hand, D. J. & Till, R. J. A simple generalisation of the area under the ROC curve for multiple class classification problems. *Machine learning* **45**, 171-186 (2001).

15 Nandy, A. S., Nassi, J. J. & Reynolds, J. H. Laminar organization of attentional modulation in macaque visual area V4. *Neuron* **93**, 235-246 (2017).

16 Nandy, A., Nassi, J. J., Jadi, M. P. & Reynolds, J. Optogenetically induced low-frequency correlations impair perception. *Elife* **8**, e35123 (2019).

17 Mitzdorf, U. Current source-density method and application in cat cerebral cortex: investigation of evoked potentials and EEG phenomena. *Physiol Rev* **65**, 37-100 (1985).

Figure 1

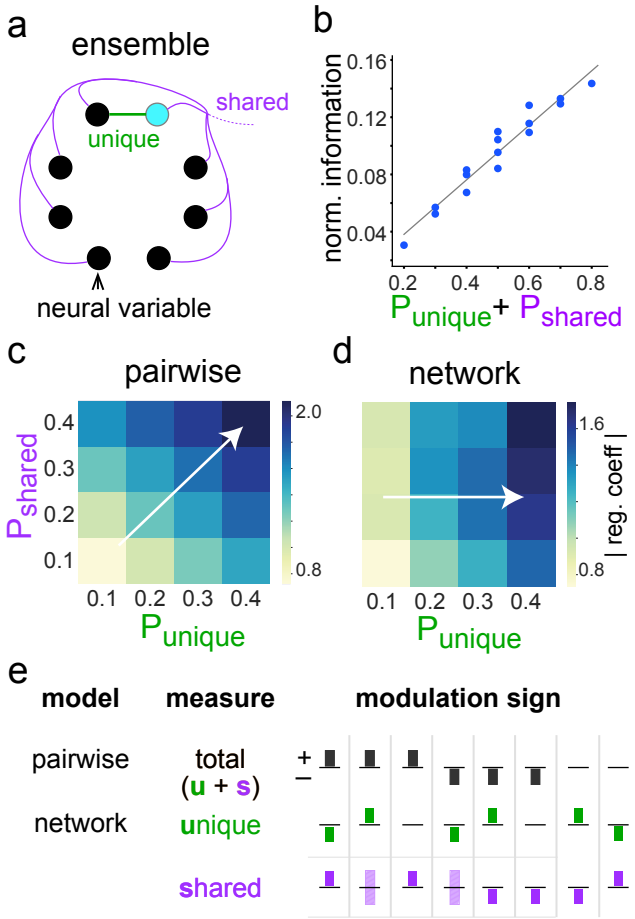


**Figure 1 Pairwise- and network-model parameters as a function of unique vs. shared dependencies. a** A synthetic ensemble of eight neural variables with two kinds of dependencies – unique or shared – between seven source variables (black) and one target variable (cyan). All interactions are excitatory. Strength of dependencies is determined by model parameters P_unique and P_shared (see SI Methods). **b** Normalized total mutual information, measured by uncertainty coefficient, as a function of the sum of model parameters (P_unique,P_shared ) that varied unique and shared components of mutual information in a monotonic way (Fig S1c-d). **c** Coefficients of pairwise model (univariate logistic regression) as a function of P_unique and P_shared. White arrow provides a visual guide for direction of highest change in coefficients. **d** Coefficients of network model (multivariate LASSO regression) as a function of P_unique and P_shared. White arrow provides a visual guide for direction of highest change in coefficients. **e** Schema for utilizing pairwise and network methods for the estimation of total (black) and unique (green) information modulation respectively, and to infer the modulation of shared information (purple). Shaded blocks indicate indeterminate modulation direction of shared information in the network.

Figure 2

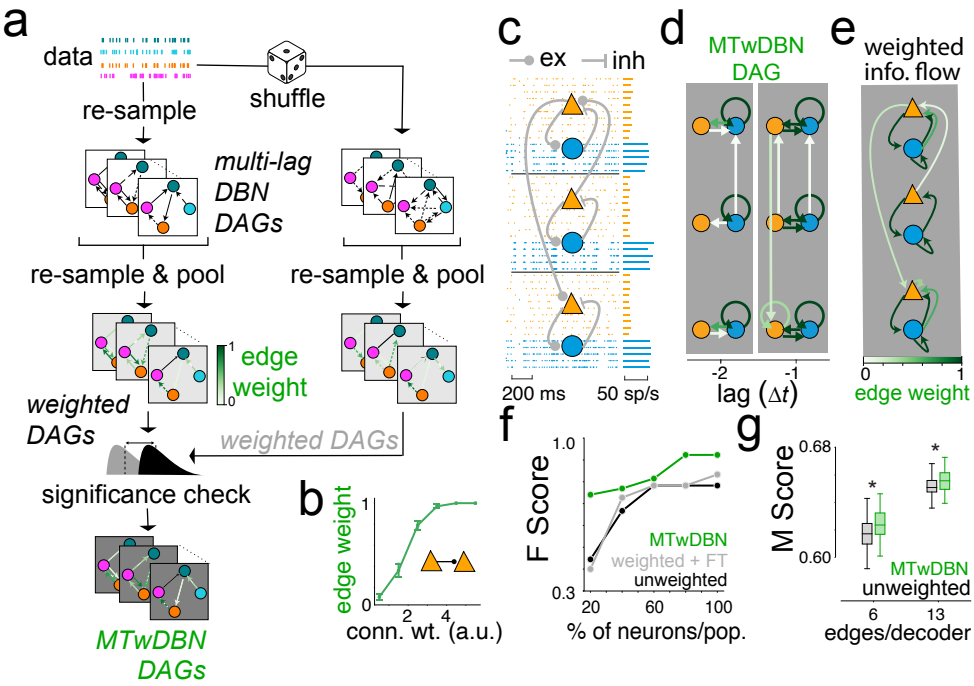


**Figure 2 DBN based estimation of multi-timescale weighted unique dependency. a** Analysis flow for multi-timescale weighted DBN (MTwDBN) model fitting. **b** Edge weight of MTwDBN as a function of connection weight in a 2-population simulated network using the pipeline in a. **c** Spiking activity of 6 subpopulations in a simulated laminar network. Connectivity is visualized in the overlaid schematic. **d** Directed dependencies (edge in the graph) in the simulated network in **c**, estimated using MTwDBN fitting. **e** Summary graph of dependencies across all timescales from **d**. **f** F score (harmonic mean of precession and recall of dependency structure) as a function of % of population observed. F score was estimated for shuffle corrected weighted DAGs (MTwDBN), weighted DAGs with fixed threshold (FT), or unweighted DAGs. **g** DBN decoder accuracy with different sizes of MTwDBN DAGs. Graph edges for the decoder were sampled from the learned structure either in an unbiased fashion or biased with the edge weights.

Figure 3

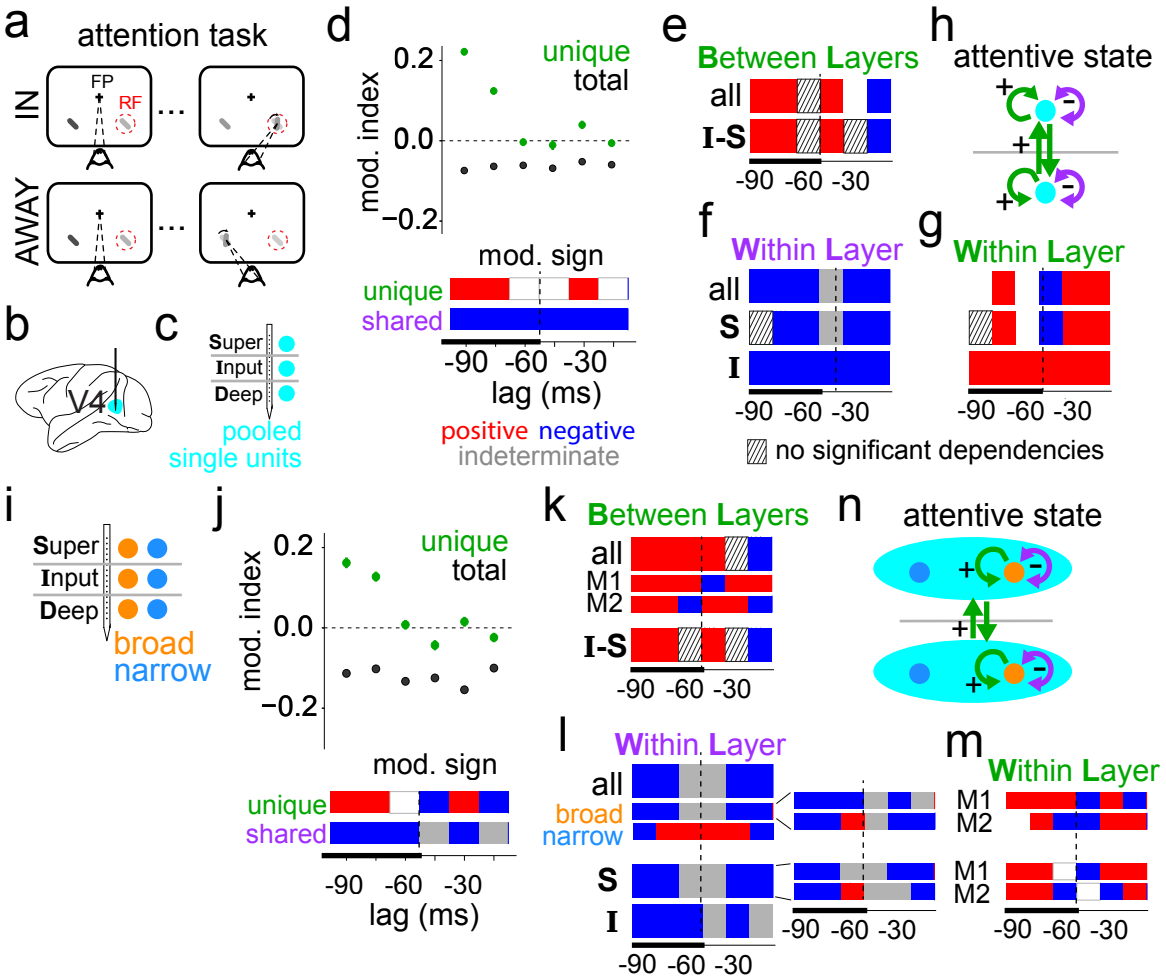


**Figure 3 Modulation of dependencies in a V4 laminar network across attention conditions.** **a** Experimental protocol: Paired Gabor stimuli with different contrasts (see Methods); one stimulus was presented inside the receptive fields (RFs) of the recorded neurons and the other at an equally eccentric location across the vertical meridian. Attention was cued either to the neurons' RFs ("IN") or to the location in the contralateral visual hemifield ("AWAY"). Orientation of one of the two stimuli changed at a random time. Monkey was rewarded for detecting the change by making a saccade to the corresponding location. Task difficulty was controlled by the magnitude of orientation change. **b** Recording approach: laminar recordings in visual area V4. **c** Neural populations used for MTwDBN analysis. Current source density analysis identified different layers (superficial, input, deep), and isolated single units were assigned to one of these layers (see SI Methods). **d** Top: MTwDBN-based modulation (green) of all unique dependencies between the laminar populations. Modulation of the same dependencies as estimated by logistic regression (black). Bottom: PID framework-based estimated modulation sign of shared dependencies (see schema in Fig 1e). Thicker line along the time axis indicates the timescales of attentional modulation in prior studies34,35. **e** Sign of average modulation of unique dependencies between layers. **f** Sign of average modulation of shared dependencies within layers. **g** Sign of average modulation of unique dependencies within layers. **h** Summary of dependency modulation pattern. **i** Neural populations used for laminar MTwDBN analysis. Isolated single units were classified as broad and narrow based on peak-to-trough duration in their average spike shape (see SI Methods). **j-n** Same as **d-h**, but for populations shown in **i.** I: input layer; S: superficial layer; M1, M2: subject-wise dependencies; broad, narrow: cell-class specific dependencies. See Fig S4 for modulation indices in **e-g**, **k-n**.

Figure 4

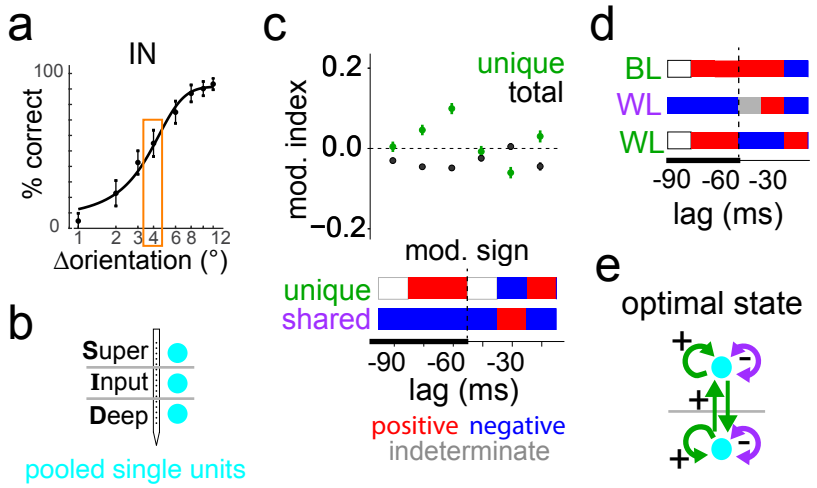


**Figure 4 Modulation of dependencies in a V4 laminar network across behavioral outcomes at perceptual threshold. a** Example session showing performance as a function of task difficulty. (orange box: threshold orientation change at which the animal was equally likely to correctly detect (Hit) or fail to detect (Miss) the change). **b** Laminar populations used for multi-lag analysis. **c** Modulation magnitude (top) and sign (bottom) of all unique (green) and total (black) laminar dependencies in **b** across Hits and Misses at perceptual threshold. Estimated modulation sign of shared dependencies (bottom, see Fig 1e). **d** Modulation sign of between (BL) and within (WL) layer dependencies. See Fig S5 for modulation indices. **e** Summary of laminar dependency modulation pattern.

Figure S1

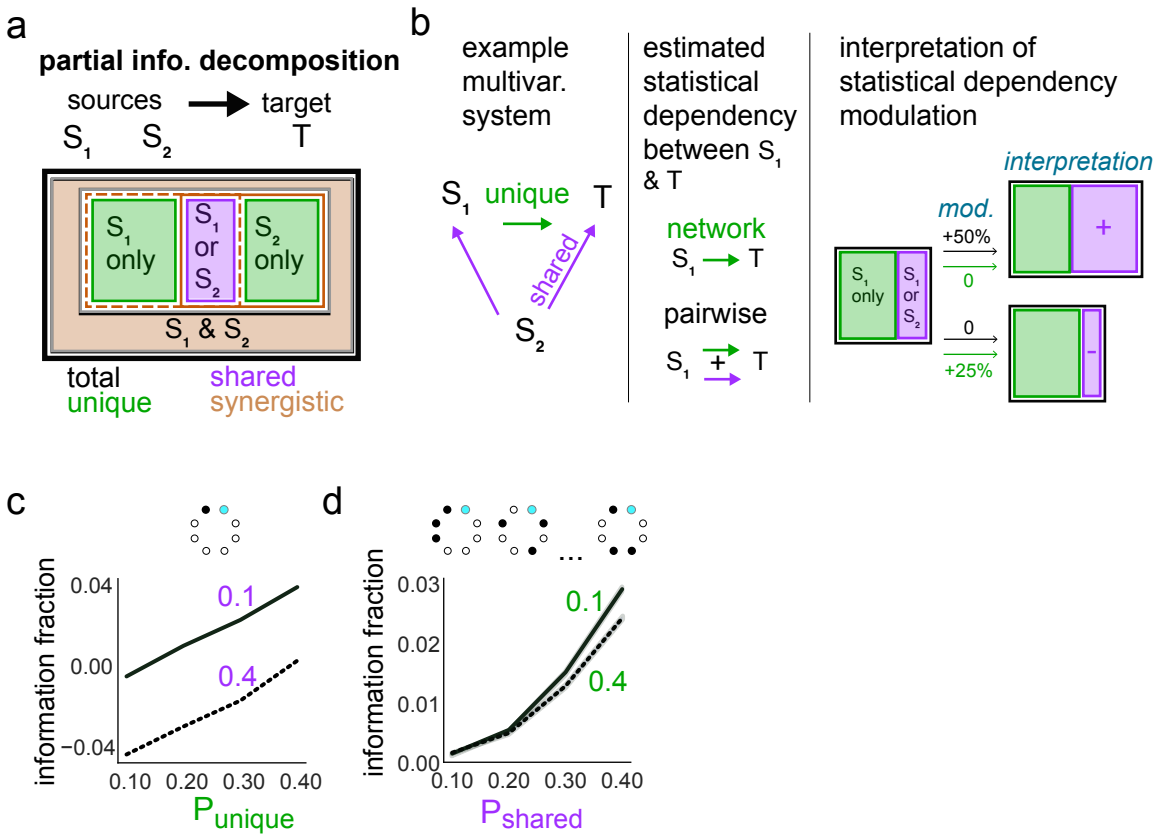


**Figure S1 Information decomposition in a multi-variate system**

**a** Partial information decomposition (PID) framework based definition of types of information that multiple sources can have about a target.

**b** Application of pairwise and network-based statistical models for approximate information decomposition in a multivariate system. Panel on the right illustrates interpretation of the modulation of these dependencies using the PID framework.

**c, d** Reduction in the proportion of total entropy (or information fraction) as a function of parameters (P_unique,P_shared) that control unique and shared information in the model. For the sake of computational efficiency, information fraction estimation as a function of Pshared (d) was performed using a subset of variables in the simulated network.

Figure S2

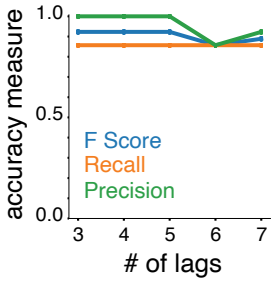


**Figure S2 Additional MTwDBN performance metric for synthetic data**
Recall, precision and F score measures for weighted DAGs in Fig 2d as a function of number of lags.

Figure S3

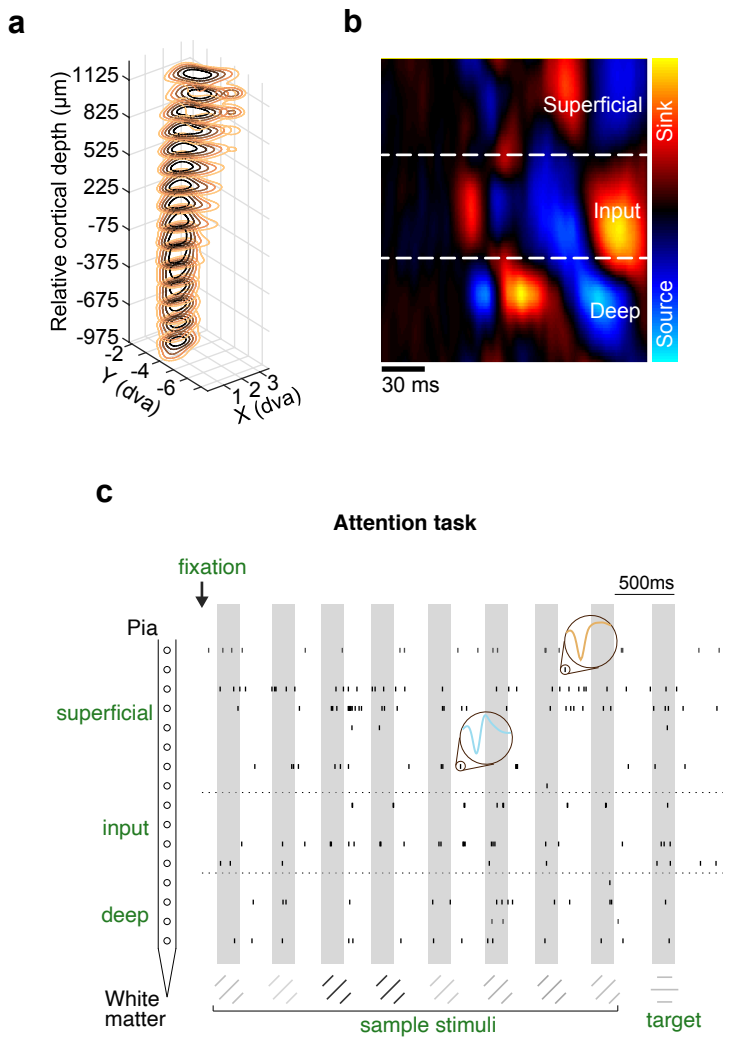


**Figure S3 Laminar recordings in area V4**

**a** Stacked contour plot showing spatial receptive fields (RFs) along the laminar probe from an example session. The RFs are well aligned, indicating perpendicular penetration down a cortical column. Zero depth represents the center of the input layer as estimated with current source density (CSD) analysis. **b** CSD displayed as a colored map. The x-axis represents time from stimulus onset; the y-axis represents cortical depth. The CSD map has been spatially smoothed for visualization. **c** An example trial showing single-unit activity across the cortical depth in the attend-in condition. The time axis is referenced to the appearance of the fixation spot. Spikes (vertical ticks) in each recording channel come from the single unit with the highest spike rate in this trial. The gray boxes depict stimulus presentation epochs. In this particular trial, 8 sample stimuli with different contrasts were presented, followed by a target stimulus flash with an orientation change that the monkey responded to correctly. Spike waveforms for two example narrow (cyan) and broad (orange) single units are shown.

Figure S4

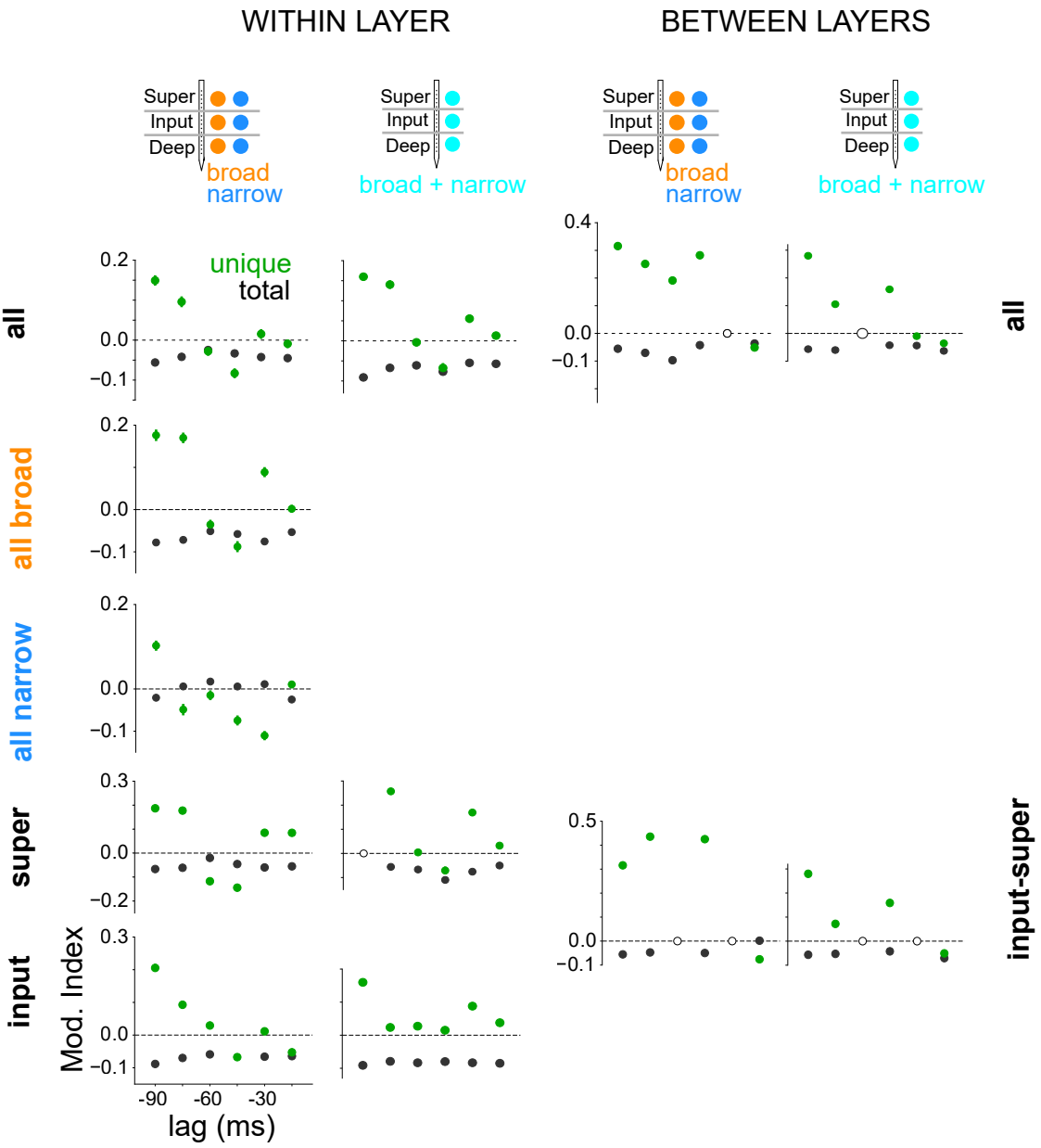


**Figure S4 Modulation of dependencies in a V4 laminar network across attention conditions.** MTwDBN-based modulation (green) of all unique dependencies between the laminar populations. Modulation of the same dependencies as estimated by logistic regression (black). Open circles indicate no significant edges for a given time lag. Schematic at the top of every column depicts the populations used for identifying laminar dependency structure within (left) and between (right) cortical layers. broad, narrow: cell-class specific populations identified by spike shape (see SI Methods).

Figure S5

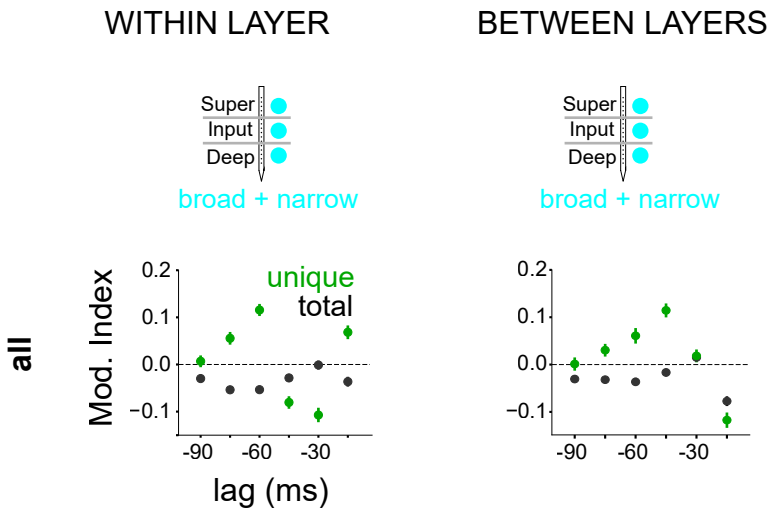


**Figure S5 Modulation of dependencies in a V4 laminar network across behavioral outcomes at perceptual threshold.** MTwDBN-based modulation (green) of all unique dependencies between the laminar populations. Modulation of the same dependencies as estimated by logistic regression (black). Schematic at the top of every column depicts the populations used for identifying laminar dependency structure within (left) and between (right) cortical layers.