Brain-state mediated modulation of inter-laminar dependencies in visual cortex

Anirban Das[1,2,7,#], Alec G. Sheffield[4,#], Anirvan S. Nandy[2,3,5,6,†] and Monika P. Jadi[1,2,5,†,*]

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

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

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

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

[5]Wu Tsai Institute, Yale University, New Haven, CT 06511

[6]Kavli Institute for Neuroscience, Yale University, New Haven, CT 06511

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

[#]equal contribution

[†]senior author

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

ABSTRACT

Spatial attention is a quintessential example of adaptive information processing in the brain and is critical for recognizing behaviorally relevant objects in a cluttered environment. Object recognition is mediated by neural encoding along the ventral visual hierarchy. How the deployment of spatial attention aids these hierarchical computations is unclear. Prior studies point to two distinct mechanisms: an improvement in the efficacy of information directed from one encoding stage to another[1-5], and/or a suppression of shared information within encoding stages[6,7]. To test these proposals, it is crucial to estimate the attentional modulation of unique information flow across and shared information within the encoding stages of the visual hierarchy. We investigated this in the multi-stage laminar network of visual area V4, an area strongly modulated by attention[8-10]. Using network-based dependency estimation from multivariate data, we quantified the modulation of inter-layer information flow during 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[11,12], we estimated the modulation of shared dependencies and found that they are reduced specifically in the putative excitatory subpopulations within a layer. Surprisingly, we found a strengthening of unique dependencies within the laminar populations, a finding not previously predicted. Crucially, these modulation patterns were also observed during successful behavioral outcomes (hits) that are thought to be mediated by endogenous brain state fluctuations[13-15], and not by experimentally imposed attentive states. Finally, phases of endogenous fluctuations that were optimal for 'hits' were associated with reduced neural excitability. A reduction in neural excitability, potentially mediated by diminished shared inputs, suggests a novel mechanism for enhancing unique information transmission during optimal states. By decomposing the modulation of multivariate information, and combined with prior theoretical work[16], our results suggest common computations of optimal sensory states that are attained by either task demands or endogenous fluctuations.

INTRODUCTION

Adaptive information processing, comprised of local computations and their efficient routing, is crucial for flexible brain function. Attention to features, such as color and shape, or locations of interest in the visual scene is regularly deployed by the visual system to enhance object recognition[17,18]. Object recognition is mediated by neural computations distributed across the ventral visual hierarchy in the cortex - V1, V2, V4, and IT – with increasing receptive field sizes and complexity of features encoded along the way[19,20]. Deployment of spatial attention is thought to modulate these hierarchical computations in a way that aids both information representation and information transfer. Neurons that encode an attended visual stimulus increase their activity at various stages (V1, V2, V4, LIP, MT) of visual processing[11,21-27,10,28], with neuronal gain

especially prominent in the later stages of the hierarchy. Attentional gain increase is also cell-class[29-33] and layer specific in area V4[7]. Based on pairwise correlation analysis, two additional mechanisms of attentional enhancement 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[1-5]. Based on theoretical results that even weak correlated variability can substantially limit the encoding capacity of a neuronal pool[16], a second proposal is an improvement in the sensory information capacity of an encoding stage through a reduction in shared fluctuations[6,7] of neural activity. However, pairwise analyses capture both unique and shared components of these fluctuations, and therefore cannot disambiguate the proposed mechanisms in a multi-variate system such as the cortical network. 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 these questions in the multi-stage laminar network of visual area V4, an area in the ventral visual stream that is strongly modulated by attention[8-10]. The visual cortex has a laminar organization and both sensory computations and information flow patterns are layer-specific, forming the building blocks of the ventral visual hierarchy[34-38]. We hypothesized that the deployment of spatial attention strengthens unique inter-layer information transfer between and weakens shared information within the input and superficial layers, both of which are crucial nodes of feedforward information flow along the ventral visual hierarchy[35,39]. To test this, our primary goal was to quantify unique statistical dependencies between the populations of each layer, which requires characterizing the joint spiking activity of the laminar ensemble. Using network-based statistical modeling, we estimated the strength of inter-layer information flow by measuring statistical dependencies in the V4 network 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. Using the partial information decomposition framework[11,12], we estimated the

76 modulation of shared dependencies, specifically in the putative excitatory subpopulations.
77 Additionally, we assessed if optimal sensory processing strategies are common to brain states
78 that could be either task-driven or resulting from endogenous state fluctuations[13-15].

79 **RESULTS**

80 **Information decomposition framework to distinguish the modulation of unique and shared**
81 **sources of dependencies in a network**

82 In a multi-actor system, it is important to understand how information carried by multiple source
83 variables about a target variable is distributed over the source variables. Partial information
84 decomposition[11,12] (PID) seeks to characterize the multivariate Shannon information that a set of
85 source variables contains about a target variable into basic 'atoms of information'[40]. Depending
86 on the nature of the underlying interactions between actors, multivariate information can be
87 decomposed into two or more key components: uniquely directed from each source, shared by
88 sources (Fig 1a), or additionally, synergistically provided by two or more sources (Fig S1). Cortical
89 neural ensembles are highly interconnected, and this can be a source of both unique and shared
90 dependencies between any pair of ensemble members (Fig 1a-b). We hypothesized that
91 dependencies traditionally captured by pairwise models should reflect the total information that a
92 source has about a target, while those captured by network models should mainly reflect the
93 unique information[41,42]. We tested this in synthetic multivariate neural data (see Methods) with
94 parameterized unique and shared dependencies between variables (Fig 1b-c). When we fit a
95 univariate regression model to our data, the coefficients indeed varied as a function of the sum of
96 shared and unique information (Fig 1c,d). On the other hand, when a multivariate LASSO
97 regression[43] was fit to our synthetic data, the regression coefficients were highly sensitive to the
98 unique information and weakly so to the shared information (Fig 1c,e). In summary, pairwise
99 measures reflect both unique and shared sources of dependencies, whereas network-based

100 measures that penalize correlations between source variables (such as regularized regression
101 coefficients), mainly extract unique sources of dependencies. With a goal to identify the
102 modulation of unique and shared dependencies across experimental conditions, we proceeded
103 to use a combination of network-based and pairwise models, interpreted within the PID
104 framework, to analyze the neural spiking data and "decompose" the modulation of unique and
105 shared dependency components by spatial attention (Fig 1f,g). Since neural dependencies are
106 temporally directed and can manifest at multiple timescales, we first developed a network-based
107 measure of Granger[44] causal dependencies that discovers unique information that a neural
108 source has about its target in an unbiased manner.

109

110 **Unbiased estimation of Granger causal dependencies**

111 Probabilistic graphical models (PGM)[45] capture the factorization of the joint distribution of a large
112 number of random variables that interact with each other. The graph edges express the
113 conditional dependence structure between the variables that form its nodes (Fig 2a). These
114 models provide a framework for capturing the unique dependency structures among multiple
115 variables, without explicit assumptions about the nature of dependency. Dynamic Bayesian
116 Networks (DBN)[46-51] is a class of PGMs wherein graph edges describe Granger causal
117 dependencies between multiple variables in a sparse structure. The graph edges express the
118 conditional dependence structure between the variables and their lags (Fig 2b). Fitting our data
119 with DBNs allows unbiased discovery of multi-timescale Granger causal[44] and cyclical
120 dependencies, without making assumptions about their specific nature (linear/nonlinear),
121 direction, or latency (Fig 2c, Fig S2). In addition to fitting models with multiple lags, we estimated
122 the probability of discovered dependencies (referred to henceforth as weights of a graph edge)
123 and used it as a measure of their strength[52] (Fig 2c,d). We used a time-shuffled estimate of the
124 edge weights to determine the statistical significance of the discovered dependencies (Fig 2c).

125 We refer to this approach as multi-timescale weighted Dynamic Bayesian Networks (MTwDBN).
126 To validate our analysis pipeline, we simulated a three-layer (Superficial, Input, Deep) cortical
127 network of stochastically spiking excitatory and inhibitory neurons that were recurrently connected
128 within layers, but followed the anatomically informed dominant patterns of interlaminar excitatory
129 connectivity (Fig 2e). When tested on synthetic data based on spiking activity in the network, the
130 MTwDBN approach robustly recovered the dependency structure (Fig 2e-g), and performed
131 significantly better than regularized regression, especially when the populations were sparsely
132 sampled (Fig 2h) as is the case in neural recordings[53-55]. Additionally, MTwDBN outperformed
133 existing unweighted[51] or fixed-threshold weighted DBN methods[52] in uncovering the dependency
134 structure (Fig 2h). Finally, in addition to dependency strength, edge weights in MTwDBN provided
135 a more accurate estimate of the dependency structure (Fig 2i). Based on these findings, we
136 proceeded to analyze neural data within the PID framework, using the modulation of MTwDBN
137 edge weights as a measure of the changes in unique dependencies across attention conditions
138 (Fig 2j).
139

140 **Attentional modulation of unique and shared dependencies in V4 ensembles**

141 Visual area V4 is strongly modulated by attention[6,9,30,31,56], with neurons exhibiting attention-
142 mediated gain increases in a cell-class[29-33] and layer-specific manner[7]. However, it is not known
143 whether and how attention modulates information flow in the laminar cortical circuit. To
144 characterize this, we employed linear array electrodes and recorded neuronal activity from well-
145 isolated single units simultaneously across the cortical depth in visual area V4 of macaque
146 monkeys performing an attention-demanding orientation change detection task (Fig 3, see
147 Methods; 2 animals, 337 single units). In the main experiment, we presented a variable-length
148 sequence of paired Gabor stimuli with different contrasts[30] (Fig 3a), with one stimulus overlapping

149 the receptive fields (RF) of the recorded sites. Attention was cued either to the stimulus within the
150 neurons' receptive fields ("IN") or to the one outside it ("AWAY"). The monkey was rewarded for
151 detecting the change by making a saccade to one of the two stimuli that changed its orientation
152 at a time unknown to the animal. We used current source density (CSD) analysis to identify
153 different laminar compartments (*superficial*, *input*, and *deep*), and assigned isolated single units
154 to one of the three layers[30] (Fig 3b, see Methods). To characterize the modulation of information
155 flow, we analyzed layer-wise pooled spiking activity of isolated single units only, as they could be
156 classified based on their average spike shape (Fig 3c, see Methods).
157
158 We fit MTwDBN and logistic regression models to the pooled spiking activity of neural
159 subpopulations defined by layers and neuronal types (Fig 4). We first considered the ensemble
160 of layer-wise populations (Fig 4a) and quantified the net attentional modulation of dependencies
161 across multiple timescales (Fig 4b). At longer timescales (> 60 ms lag)[33], while attention
162 weakened pairwise dependencies estimated by the regression models, in agreement with
163 previous findings[57,58], we found a strengthening of unique dependencies overall. The inferred
164 modulation of shared component of dependencies showed a weakening by attention (Fig 4b,
165 bottom), thereby providing direct evidence for this previously hypothesized mechanism of
166 perceptual improvement by attention[57,58]. MTwDBN-based unique dependencies between layers,
167 specifically those between the input and superficial layers, an important link in feedforward
168 processing, were strengthened by attention (Fig 4c). On the other hand, shared dependencies
169 within the input and superficial layers were inferred to be weakened by attention (Fig 4d).
170 Surprisingly, unique dependencies within the layers were strengthened by attention at most
171 timescales (Fig 4e). Taken together, these results provide the first direct evidence that attention
172 improves unique dependencies both within and across stages of the laminar circuits of the ventral
173 visual hierarchy, while weakening shared dependencies within these stages (Fig 4f).

174

175 Neuronal cell classes in the cortex contribute differentially to information processing[30,31,59]. To test
176 if the above results hold when we allow the discovery of cell-class specific dependencies, we next
177 analyzed an ensemble of broad- and narrow-spiking layer-wise populations (Fig 4g) and
178 quantified the net attentional modulation of dependencies across different timescales (Fig 4h).
179 The pattern of net modulation of unique and shared dependencies in this ensemble largely
180 mirrored that discovered in the layer-wise aggregated ensemble, specifically at the longer
181 timescales. Same was the case for the modulation of unique dependencies between layers and
182 shared dependencies within layers (consistent across both animals) (Fig 4i,j). Theoretical
183 predictions[16] regarding the effects of shared spike-count covariation on information representation
184 apply primarily to excitatory neurons which are the primary projection neurons in the cortex. When
185 we quantified the attention modulation of shared dependencies within layers in a cell-class
186 specific manner, we found that the broad-spiking population (putative excitatory neurons) showed
187 a robust weakening of shared dependencies within layers (consistent across subjects) (Fig 4j).
188 On the other hand, the narrow-spiking population (putative inhibitory neurons) showed a distinct
189 pattern, one that was dominated by a strengthening of shared dependencies within layers.
190 Consistent with the findings in the aggregated ensemble, unique dependencies within the layers
191 were strengthened by attention at most timescales (Fig 4k). Taken together, these results provide
192 the first direct evidence that attention specifically weakens shared dependencies in the projection
193 populations within encoding stages, in addition to robustly improving unique dependencies both
194 within and across stages of the ventral visual hierarchy (Fig 4l).

195

196 **Modulation of unique and shared dependencies in V4 ensembles by behavior outcome**

197 To test if the pattern of inter- and intra-layer dependency modulation that we observed is a
198 signature of brain states that are optimal for perceptual behavior, we analyzed the laminar

199 ensemble activity for a subset of trials within the attend-in condition in which the animal was
200 equally likely to correctly detect (Hit) or fail to detect (Miss) the orientation change (Fig 5a,b).
201 Controlling for task and stimulus conditions, these behavioral fluctuations are thought to arise
202 from endogenous brain state fluctuations such as attentional sampling and arousal changes[13-
203 15,60-64]. The pattern of net modulation of unique and shared dependencies across behavioral
204 outcomes (Fig 5c) largely mirrored the pattern discovered across attentive states (Fig 5d) at lags
205 longer than 60 ms. The same was the case for modulation of unique dependencies between
206 layers and shared dependencies within layers (Fig 5d), suggesting that the pattern of enhanced
207 inter-layer unique dependencies and weakened intra-layer shared dependencies is a hallmark of
208 brain states that are optimal for perceptual performance (Fig 5e). Combined with prior theoretical
209 work[16], our results suggest a conceptual model of optimal states that involves enhanced inter-
210 layer communication and improved intra-layer information capacity that are either imposed by
211 task demands or are attained via endogenous fluctuations in brain state.
212
213 The above model hypothesizes that endogenous fluctuations modulate the hit rate through
214 periods of lower shared variability. To test if optimal states that are associated with hit trials are
215 also associated with a reduction in shared correlations among the projection neurons of the input
216 layer, we estimated the probability of presentation of "successful" target stimuli and the probability
217 of spiking of input layer broad spiking units, both as a function of the phase of the ongoing cortical
218 activity. We estimated the generalized phase of the band-filtered (5-40 Hz) local field potential
219 signals in the input layer (see Methods), and calculated the probability of a 'hit'-causing target
220 onset and of neuronal spikes at different phases (Fig 5f). We found a clear phase dependence of
221 response onset of 'hit' targets (Fig 5g, top). These phases were also associated with a lower
222 excitability of broad spiking cells (Fig 5g, bottom), suggesting that the improved performance in
223 optimal states occurs during phases of lower than average spiking probability of putative

224 excitatory neurons in the input layer. Interestingly, the excitability of superficial layer putative
225 excitatory neurons, the primary candidates that project to downstream cortical areas in the ventral
226 stream, was independent of the phase of the ongoing activity in the superficial layers of V4 (Fig
227 S5).
228
229
230 **DISCUSSION**
231 The laminar network is considered a canonical circuit that constitutes a key computational unit in
232 the cortex. While anatomical connectivity maps have identified the key variables of this unit[35,39],
233 access to the functional connectivity that determines the resulting computations has remained
234 elusive. Laminar recordings in awake animals transitioning across behavioral states have
235 allowed us to observe the neural variables that are expected to play a significant role in these
236 computations. Using a novel combination of network-based dependency models and information
237 decomposition, we show that the dependency structure in the laminar cortical ensemble is
238 modulated by attention in a layer and cell-type specific manner. The input and superficial layers
239 in V4 are crucial nodes of feedforward information flow along the ventral visual hierarchy, with the
240 input layer receiving information from earlier visual areas (V1, V2) and excitatory projection
241 neurons in the superficial layer sending information to the next stage of this hierarchy, namely the
242 inferotemporal cortex[35,39]. We find that unique dependencies between input and superficial layer
243 populations are strengthened by attention, as well as during successful behavioral outcomes
244 within the attentive state. This finding suggests that enhanced unique information transfer
245 between encoding stages of the laminar hierarchy is a hallmark of behaviorally optimal sensory
246 processing. On the other hand, shared dependencies within the putative excitatory laminar
247 populations are weakened by attention, as well as during successful behavioral outcomes. Prior
248 theoretical and modeling studies have proposed that a reduction in shared correlations can

249 enhance both the information capacity[16] and signal-to-noise ratio[7,57] of a neural population.
250 Interpreted within this framework, our finding suggests that enhanced information capacity
251 mediated by reduced shared dependencies is another hallmark of behaviorally optimal sensory
252 processing.
253
254 An unexpected finding of our study is the attentional enhancement of unique dependencies within
255 the (putative) excitatory laminar populations. Prior theoretical work has shown that networks with
256 clustered (as opposed to random) architectures can result in dynamics that lead to strong
257 correlations within the clustered subpopulations[65]. Excitatory neurons within the superficial layers
258 (layers 2/3) in the visual cortex have been shown to exhibit higher connection probability when
259 the neurons receive common inputs from the input layer (layer 4) and superficial layers[66]. Since
260 our laminar populations are tightly localized in cortical space and, given their strongly overlapping
261 receptive fields (Fig. 3b), are recipients of highly overlapping inputs, their activity fluctuations
262 could result in within-population unique dependencies. The observed modulations of these
263 dependencies could reflect the attentional enhancement of effective connectivity (and hence
264 clustering) in these populations and the resulting enhancement in neural activity fluctuations.
265
266 Additionally, we find that phases of the endogenous fluctuations that are associated with optimal
267 target presentation (resulting in hits) are also associated with reduced excitability of broad spiking
268 neurons, especially in the input layer. Our findings are in agreement with previous reports of
269 rhythmic shifts of neural excitability and their entrainment to the stream of sensory inputs as key
270 mechanisms of sensory selection[60-64,67]. Interpreting the fluctuations in excitability to be at least
271 partly based on changing correlations due to fluctuations in shared inputs, our finding suggests
272 an additional mechanism through which weakened shared neural activity fluctuations could
273 improve behavioral outcome: a lowered excitability, which is associated with reliable encoding in

the visual cortex[68].

Our findings and interpretations point to a set of key questions for future studies that will shed further light on the mechanisms of optimal states. One key question that can be addressed using our computational approach, but is currently limited by the statistical power from available datasets is: how are unique and shared dependencies between pairs of populations modulated in specific causal directions, and between specific cell classes? While the current data allowed us to answer this in a limited way, recent advances in recording techniques should allow the collection of denser neural data that, in combination with the approach we successfully demonstrated in this study, would provide further insights into the computations and mechanisms of optimal behavioral states. Another equally significant question is: How are unique and shared dependencies between superficial and deep layers of the laminar network modulated during optimal behavioral states? The communication between superficial and deep layers, specifically in the direction from the former to the latter, plays a key role in the feedback computations of the ventral visual hierarchy. Theoretical frameworks of object perception suggest computations of hierarchical Bayesian Inference, with a key role for feedback pathways[69,70]. Additionally, our group has shown previously that optogenetic induction of low-frequency fluctuations in V4 impairs performance in an attention-demanding task[71]. An important follow-up to this line of investigation would be testing the hypothesis that reduced shared dependencies in laminar populations are causal to enhanced information capacity. This will require experimental paradigms with a richer set of visual stimulation and causal manipulation of shared neural activity fluctuations in a layer-specific manner.

MTwDBN analysis approach

Our approach demonstrates that the information flow structure in a neural ensemble is robustly

299 described by Dynamic Bayesian Network modeling that is adapted to include multiple lags
300 spanning timescales of interest. This unbiased approach allows us to dive deeper into the causal
301 functional interactions (beyond correlations) in a multivariate system with unknown nonlinear
302 dependencies. Further, the results demonstrate that weighing dependencies using confidence
303 measures provides a more accurate information flow structure that can be utilized to investigate
304 how this structure is modulated by changes in brain and behavioral states. Finally, an integration
305 of this approach with pairwise models allows us to dissect changes in distinct components of
306 dependencies. Since temporally directed unique dependencies suggest causality and shared
307 ones suggest common inputs, this allows us to gain insights into the network mechanisms. In
308 comparison, prior correlational studies have quantified net dependency modulation [1,7,57]. While
309 these provided valuable characterization of the neural correlates of attention, the underlying
310 network mechanisms remained hypothetical.
311
312 ***Comparison to other multivariate models***
313 While approaches such as generalized linear models (GLM) have been useful in providing
314 improved phenomenological models of individual neural responses to sensory stimulation in early
315 sensory circuits, such as in the retina[72], they are not ideal for dependency structure learning in
316 the highly recurrent cortical ensemble[35]. It is possible to learn the dependency structure in a
317 laminar network by fitting a GLM model for each of the observed populations in our data, with the
318 remaining populations as predictors. However, the assumption of the independence of predictor
319 variables inherent in GLMs can result in the discovery of spurious dependencies due to shared
320 inputs, and hence, a dense and likely inaccurate structure. Since structure learning in DBNs
321 involves determining conditional independence by solving an optimization problem that penalizes
322 density, our approach is ideal for generating a sparse and hence interpretable unique dependency
323 structure in multivariate data.

Current limitations of our approach

Currently, a limitation of this class of DBN models is that it 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 the structure and modulation of information flow in a multivariate system such as the laminar cortical network.

In general, the discovery of dependency structure, as well as its interpretation using the PID framework, 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 of dependency modulations, suggesting the 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.

Finally, the PID framework describes a synergistic component of information in a multivariate system, such as the cortical network, that we ignore for reasons of simplification. An example of synergistic information would be that conveyed by the coactivation of two or more neural populations that are causally connected to a target population. While it is a real possibility in the laminar cortical networks, this limitation of our approach will merit revisiting when significantly denser recordings are made available through future experiments.

Conclusion

Network mechanisms of behaviorally optimal brain states are shared across those that are either externally induced or are a consequence of endogenous brain state fluctuations. They target efficiencies in both information transfer and representation. MTwDBN-based models provide a quantitative description of information flow patterns 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.

FIGURE CAPTIONS

Figure 1 Dependency decomposition in a multi-variate system using pairwise and network models.

a Simplified Partial information decomposition (PID) framework[11,12] based definition of types of information that multiple sources can have about a target (see Methods and Supp. Fig S1).

b 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 Methods).

c1, c2 Information fraction (reduction in the proportion of total entropy) as a function of

parameters (P_{unique}, P_{shared}) that control unique and shared information in the model. Information fraction estimation as a function of P_{shared} (**c2**) was performed using a subset of variables in the simulated network for computational efficiency, (see Methods).

c3 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.

d Coefficients of a pairwise model (univariate logistic regression (UR)) as a function of P_{unique} and P_{shared}. White arrow provides a visual guide for direction of highest change in coefficients.

e Coefficients of a network model (LASSO multivariate regularized regression (RR)) as a function of P_{unique} and P_{shared}. White arrow provides a visual guide for direction of highest change in coefficients.

f Application of pairwise and network-based statistical models for approximate information decomposition in an example multivariate system. It illustrates interpretation of the modulation of these dependencies using the PID framework.

g Schema for utilizing pairwise and network methods for the estimation of total (brown) and unique (green) information modulation respectively, and to infer the modulation of shared information (purple) based on the PID framework. Shaded blocks indicate indeterminate modulation direction of shared information in the network.

Figure 2 Multi-timescale & weighted Dynamic Bayesian Networks (DBN) based estimation of unique dependencies in a sparsely sampled recurrent neuronal network. a Bayesian Networks for graph representation of dependencies in a multivariate system. **b** Dynamic Bayesian Networks (DBN) for graph representation of dependencies in multivariate time series data. **c** Analysis flow for multi-timescale weighted DBN (MTwDBN) graphical model fitting. **d** Edge weight

396 of MTwDBN graph as a function of connection weight in a 2-population simulated network using
397 the pipeline in **c**. **e** Spiking activity of 6 subpopulations in a simulated laminar network.
398 Connectivity is visualized in the overlaid schematic. **f** Directed dependencies (edge in the graph)
399 in the simulated network in **e**, estimated using MTwDBN fitting. **g** Summary graph of
400 dependencies across all timescales from **f**. Solid and dashed lines indicate two different
401 timescales. **h** F-score (harmonic mean of precision and recall of dependency structure) as a
402 function of % of neural population observed. F-score was estimated for shuffle corrected weighted
403 DAGs (MTwDBN, green), weighted DAGs with a fixed threshold (weightedFT, blue), unweighted
404 DAGs (red), or LASSO regression, an example of regularized regression (RR) models (black).
405 Each point represents the average of five separate runs, except 100 % (single run). Error bars
406 indicate standard deviation, some error bars are smaller than symbol size. **i** DBN decoder
407 accuracy with different sizes of MTwDBN DAGs. Decoders were trained to predict population
408 activities using a subsample of shuffle-corrected edges (see Methods). Graph edges for the
409 decoder were sampled from the learned structure either in an unbiased fashion (black) or biased
410 with the edge weights (green). **j** Schema for estimating modulation of unique dependencies in a
411 network of neural populations, using MTwDBN.

412 **Figure 3 Laminar recordings in area V4. a** Experimental protocol: Paired Gabor stimuli with
413 varying contrasts (see Methods); one stimulus was presented inside the receptive fields (RFs) of
414 the recorded neurons and the other at an equally eccentric location across the vertical meridian.
415 Attention was cued either to the neurons' RFs ("IN") or to the location in the contralateral visual
416 hemifield ("AWAY"). The orientation of one of the two stimuli changed at a random time. The
417 monkey was rewarded for detecting the change by making a saccade to the corresponding
418 location. Task difficulty was controlled by the magnitude of orientation change. **b** *Left,* Recording
419 approach: Laminar recordings in visual area V4. *Middle,* Stacked contour plot showing spatial

420 receptive fields (RFs) along the laminar probe from an example session. Alignment of RFs
421 indicates perpendicular penetration down a cortical column. Zero depth represents the center of
422 the input layer as estimated with current source density (CSD) analysis. *Right,* CSD is displayed
423 as a colored map. The x-axis represents time from stimulus onset; the y-axis represents cortical
424 depth. The CSD map has been spatially smoothed for visualization. **c** An example trial showing
425 single-unit activity across the cortical depth in the attend-in condition. The time axis is referenced
426 to the appearance of the fixation spot. Spikes (vertical ticks) in each recording channel come from
427 either single units (blue, orange) or multi-units (black). Spike waveforms for an example narrow-
428 spiking (blue) and a broad-spiking (orange) single unit are shown. The bars at the bottom depict
429 stimulus presentation epochs, with height indicating relative stimulus contrast.

430 **Figure 4 Modulation of dependencies in a V4 laminar network across attention conditions**.
431 **a** Neural populations used for fitting laminar MTwDBN. Current source density analysis identified
432 different layers (*superficial*, *input*, *deep*), and isolated single units were assigned to one of these
433 layers (see Methods). **b** *Top*: Average MTwDBN-based modulation (green) of all unique
434 dependencies between the laminar populations. Modulation of the same dependencies as
435 estimated by logistic regression (brown). Error bars indicate 95% confidence intervals. *Bottom*:
436 Visualization of modulation sign of unique dependencies at different lags. Combining the
437 modulation sign of unique and total dependencies (see *Top*), PID framework-based estimated
438 modulation sign of shared dependencies (using schema in Fig 1g) is also shown for different lags.
439 Thicker line along the time axis indicates the timescales of attentional modulation in prior
440 studies[38,39]. **c** Sign of average modulation of unique dependencies between layers (bi-
441 directionally). $\mathbf{I}$: input layer; $\mathbf{S}$: superficial layer. **d** Sign of average modulation of shared
442 dependencies within layers. **e** Sign of average modulation of unique dependencies within layers.
443 **f** Summary of dependency modulation pattern. **g** Neural populations used for fitting laminar

444 MTwDBN. Isolated single units were classified as broad- and narrow-spiking based on peak-to-
445 trough duration in their average spike shape (see Methods). **h** *Top*: Average MTwDBN-based
446 modulation (green) of all unique dependencies between the cell-type specific laminar populations.
447 Modulation of the same dependencies as estimated by logistic regression (brown). *Bottom*:
448 Visualization of modulation sign of unique dependencies and PID framework-based estimated
449 modulation sign of shared dependencies is also shown for different lags. **i** Sign of average
450 modulation of unique dependencies between layers. **j** Sign of average modulation of shared
451 dependencies within layers for all, broad or narrow populations. *M1*, *M2*: subject-wise; *broad*,
452 *narrow*: cell-class specific. **k** Sign of average modulation of unique dependencies within layers. **l**
453 Summary of dependency modulation pattern. See Fig S3 for modulation indices in **c-e**, **i-k**.

454 **Figure 5 Modulation of dependencies in a V4 laminar network across behavioral outcomes**
455 **at perceptual threshold**. **a** Example session showing performance as a function of task difficulty.
456 Gray box: threshold orientation change at which the animal was equally likely to correctly detect
457 (hit) or fail to detect (miss) the change. **b** Laminar populations used for multi-lag analysis. **c**
458 Modulation magnitude (top) and sign (bottom) of all unique (green) and total (brown) laminar
459 dependencies in **b** across Hits and Misses at perceptual threshold. Estimated modulation sign of
460 shared dependencies (bottom, see Fig 1g). **d** Modulation sign of between layer (BL) and within
461 layer (WL) dependencies. See Fig S4 for modulation indices. **e** Summary of laminar dependency
462 modulation pattern. **f** Wideband (5-40Hz) LFP signals (colored lines) overlaid on the raw LFP (0-
463 200Hz) signals (gray) in the input layer in a portion of an example session. The generalized phase
464 (color-coded) depicts the dominant phase of the wideband LFP. Vertical ticks indicate single-unit
465 spikes in the corresponding channel (see Methods). **g** *Top*: Target stimulus presentation
466 probability as a function of the generalized phase of the LFP (adjusted for cortical delay),
467 separated by HIT and MISS trials. Asterisk (*) indicates phases with significant differences

468 between the two trial types (Ranked sum test, corrected for multiple comparisons). *Bottom*: Spike
469 probability in the input layer as a function of generalized phase of the LFP, separately estimated
470 for putative excitatory (broad) and putative inhibitory (narrow) units. For other layers, see Fig S5.
471 Error bands indicate standard error of the mean.

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

473 Complete Partial information decomposition (PID) framework[11,12] based definition of types of
474 information that multiple sources can have about a target.

475

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

477 Recall, precision and F-score measures for weighted DAGs fit to synthetic data illustrated in
478 Fig 2e as a function of number of lags.

479

480 **Figure S3 Modulation of dependencies in a V4 laminar network across attention**
481 **conditions.** MTwDBN-based modulation (green) of all unique dependencies between the laminar
482 populations. Modulation of the same dependencies as estimated by logistic regression (black).
483 Open circles indicate no significant edges for a given time lag. Schematic at the top of every
484 column depicts the populations used for identifying laminar dependency structure within (left) and
485 between (right) cortical layers. *broad, narrow*: cell-class specific populations identified by spike
486 shape (see Methods).

487

488 **Figure S4 Modulation of dependencies in a V4 laminar network across behavioral**
489 **outcomes at perceptual threshold**. MTwDBN-based modulation (green) of all unique
490 dependencies between the laminar populations. Modulation of the same dependencies as
491 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.

Figure S5 Spike probability in the superficial (top panel) and deep layer (bottom panel) as a function of generalized phase of the LFP, separately estimated for broad-spiking (orange; putative excitatory), and narrow spiking (blue; putative inhibitory) single units. Overlaid is target stimulus presentation probability as a function of generalized phase of the LFP (adjusted for cortical delay), separated by HIT (red) and MISS (grey) trials. Asterisk (*) indicates phases with significant differences between the two trial types (Ranked sum test, corrected for multiple comparisons).

AUTHOR CONTRIBUTIONS

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

ACKNOWLEDGEMENTS

This research was supported by NIH/NEI R01 EY034605, NIH R00 EY025026 to MPJ, NIH/NEI R01 EY032555, NARSAD Young Investigator Grant, Ziegler Foundation Grant, and Yale Orthwein Scholar Funds to ASN, NIH R21 MH126072 and SFARI 875855 to MPJ & 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.

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[11,12,73,74]. The information

516 $I(T;S1,S2)$ that two 'source' variables $S1$ and $S2$ hold about a third 'target' variable T decomposes
517 into four parts: (i)$U(T;S1|S2)$, the unique information that only $S1$ (out of $S1$ and $S2$) holds about
518 T; (ii) $U(T;S2|S1)$, the unique information that only $S2$ holds about T; (iii) $R(T;S1,S2)$, the
519 redundant information that both $S1$ and $S2$ hold about T; and (iv) $S(T;S1,S2)$, the synergistic
520 information about T that only arises from knowing both $S1$ and $S2$ (see Fig S1). The set of
521 quantities $\{U(T;S1\,|S2), U(T;\,S2|S1), R(T;S1,S2), S(T;S1,S2)\}$ is called a 'partial information
522 decomposition' related as follows:

523 $$I(T;S1,S2) = U(T;S1|S2) + U(T;\,S2|S1\,) + S(T;S1,S2) + R(T;S1,S2)$$

524 $$I(T;S1\,) = U(T;S1|S2) + R(T;S1\,,S2)$$

525 $$I(T;S2) = U(T;S2|S1\,) + R(T;S1\,,S2)$$

526 Thus, the Partial Information Decomposition (PID)[11,12] framework characterizes the mutual
527 information between variables by decomposing it into unique, shared, and synergistic
528 components.

529

530 In a multi-neuronal network, unique (and synergistic) components of the mutual information
531 between neural variables due to causal interactions (unique and synergistic) are captured by
532 directed statistical dependencies. On the other hand, shared components of mutual information
533 between neural variables correspond to shared neuronal inputs, and the sign of their modulation
534 can be estimated from the modulation of unique and total mutual information (Fig 1g).

535

536 *Information Decomposition with Pairwise and Network Models*

537 a. Generation of synthetic data

538 A synthetic network was generated in which unique and shared dependencies between variables
539 were controlled. Eight variables were randomly initialized to values of 0 or 1. One variable was
540 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 values of all variables were set to 1 with probability P_{shared} and unchanged with probability $1 - P_{shared}$ and 2) the values of the target and unique source were set to 1 with probability P_{unique} and unchanged 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}.

b. Information estimation

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. Additionally, unique information between the unique source variable and target was calculated for P_{unique} ∈ {0.1, 0.2, 0.3, 0.4} and P_{shared} ∈ {0.1, 0.4}.

Unique information was divided by the total entropy of the target and termed unique information fraction. Likewise, redundant or shared information about the target from random subsets of three source variables (repeated 100 times) was calculated for P_{unique} ∈ {0.1, 0.4} and P_{shared} ∈ {0.1, 0.2, 0.3, 0.4} and divided by total entropy of the target, termed shared information fraction. Unique and redundant information were calculated using the **dit** (discrete information theory) python package[75] with redundancy measured by pointwise common change in surprisal[74].

c. Pairwise model fitting

Univariate logistic regression was performed between the unique source and target variable using the statsmodel python package[76] for all values of P_{shared} and P_{unique}.

565 d. Network model fitting

566 Multivariate regression was performed between all seven source variables and the target variable

567 using the **statsmodel** python package using L1 penalty[77-79].

568

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

570 *a. Fitting Dynamic Bayesian Network models*

571 A Dynamic Bayesian Network (DBN) framework was used to learn dependencies between neural

572 populations[80]. The **pgmpy** python package[81] (https://github.com/pgmpy/pgmpy) and custom-

573 written python code was used to fit all DBN models. Each binned-and-sliced data table was first

574 bootstrapped *B* times. For each session/condition, a hill climb tabu-search with a history window

575 of 7 was performed 120 times, each from a unique random starting graph, to find a suitable fitting

576 directed acyclic graph (DAG)[45]. The Akaike information criterion (AIC) was used[82,83] as the scoring

577 metric in the tabu-search. The variables associated with the latest time slice (0 ms) were termed

578 the *effect* variables; all others were termed potential *cause* variables. The search was restricted

579 to DAGs where edges can only be incident on an *effect* variable. Edges between *effect* variables

580 were allowed as they may capture dependencies at timescales shorter than our chosen bin

581 widths, however such edges have no causal interpretation and were excluded from further

582 analysis. The resulting DAGs were termed *unweighted*, as their edges (dependencies) are

583 described in a present/absent manner. Of the 120 starting points, only the DAG with highest AIC

584 score was used for further analysis. This resulted in *B unweighted* DAGs per session/condition.

585

586 *b. Estimation of weighted DAG*

587 The *B* DAGs from each session/condition were used to estimate weighted DAGs[52]. A pool of 100

588 weighted DAGs were estimated by taking 100 bootstrap samples from the *B unweighted* DAGs

589 and averaging: the weight $\in (0,1)$ for each edge corresponds to the proportion of the *B unweighted*

590 DAGs where the edge was present. This resulted in 100 weighted DAGs for each
591 session/condition.
592
593 *c. Testing significance of DAG edges*
594 To test for significance of the discovered edges, the 100 weighted DAGs were compared to 100
595 control DAGs that were generated in the same manner as above but with time-shuffled data. The
596 binned-and-sliced data tables were shuffled as follows: for each row, permute the data from each
597 of the 7 time slices which arose from the same population. Finally, permute each column
598 (corresponding to each population/time slice combination). To test for significance, a distribution
599 of weights was generated by combining weights for a given edge across all sessions (conditions
600 treated separately). The distribution of edge weights from unshuffled data were compared to the
601 distribution of edge weights from the shuffled data using a one-sided Mann Whitney U Test. If the
602 distribution from unshuffled data had significantly higher values than from shuffled data ($p<0.05$),
603 the edge was marked as significant for that condition and said to have survived time shuffling.
604
605 *MTwDBN Validation*
606 *a. Synthetic Neural Network*
607 Synthetic neural network models were constructed using stochastic spiking neurons[84,85].
608 Individual neurons in the model were treated as coupled, continuous-time, two-state (active and
609 quiescent) Markov processes. The active state represents a neuron firing an action potential and
610 its accompanying refractory period, whereas the quiescent states represent a neuron at rest. The
611 transition probability for the i-th neuron to decay from active to quiescent state in time *dt* was
612 $P_i(\text{active} \rightarrow \text{quiescent}) = \alpha_i\, \partial(dt)$, where α_i represented the decay rate of the active state of the
613 neuron. Parameter α_i sets the upper bound on firing rate of the stochastically spiking neuron, akin
614 to a refractory period. The transition probability for the i-th neuron to change from quiescent to

active state (i.e., spike) was $P_i(\text{quiescent} \rightarrow \text{active}) = \beta_i G(S_i)\partial(dt)$ [84]. This caused the firing probability to be a function of the input, with β_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 α_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 neuron network

The model network analyzed in Fig 2d consisted of two excitatory neurons with a single synaptic connection from the first neuron to the second neuron. Ten such two-population networks were simulated 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 network

Synthetic laminar network analyzed in Fig 2e-i consisted of simulating 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 2e. 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

638 unit, and -2 for intra-laminar connections where the presynaptic unit is an I unit. 1000 trials of
639 spiking data, each 2000 ms long, were simulated using this model.
640
641 *b. Preprocessing for MTwDBN analysis*
642 Data from single units was grouped by population in each model simulation. The "multi-unit"
643 spiking activity of these populations was used for the analysis. Before aggregating activities of
644 single units into populations, *d* % (*d* being a pre-selected number, see *Effect of sub-sampling in*
645 *synthetic laminar model*) of single units in each population were dropped (not used for analysis)
646 from the laminar model data. The data from each trial were discretized into 1.2ms bins of either
647 0, 1, or 2 to denote no spikes, one spike, or multiple spikes in a time bin. The data were lagged 2
648 times to give 3 time slices (2.4, 1.2, 0 ms). The data from all trials were concatenated together to
649 generate a single data table with 6 or 18 columns (2 or 6 populations x 3 time slices). The binned
650 and discretized spiking activity of a single population in a single time slice was viewed as a
651 variable in either the pairwise regression or DBN framework. Data with structure as produced by
652 this preprocessing step are termed binned-and-sliced data tables. Alternative models with
653 between 4-7 lags were created with no improvement in accuracy (Fig S2).
654
655 *c. Analysis: Relation between synaptic weights and DAG edge weights in two-neuron*
656 *network*
657 Binned-and-sliced data tables from each two-neuron model were passed separately through the
658 MTwDBN pipeline (B = 200) to generate 100 weighted DAGs each. Weights were obtained by
659 averaging across the 200 *unweighted* DAGs. The 100 weighted DAGs were used to compute
660 95% confidence intervals for these weights.
661
662 *d. Analysis: Effect of sub-sampling on recovering ground-truth in synthetic laminar network*

To assess the effect of sub-sampling neural data on MTwDBN outputs, *d* % (d ∈ {0, 20, 40, 60, 80}) of neurons in each of 6 populations were dropped from the laminar network before preprocessing (*Preprocessing Synthetic Neural Network Data).* For *d* ∈ {20, 40, 60, 80}, five iterations of this procedure were performed where a different random subset of neurons was dropped. This resulted in 21 binned-and-sliced data tables.

1. *DBN Models: MTwDBN, Weighted with Fixed Threshold, Unweighted*

Each data table was passed through the MTwDBN pipeline (*B* = 200), generating 200 unweighted DAGs and 100 weighted DAGs. We used bootstrap sizes of 8000 rows in the MTwDBN pipeline to simulate data scarcity of electrophysiology experiments. Edges were considered significant either by surviving time shuffling (MTwDBN) or by the average weighted edge passing a fixed threshold of 0.5 (weightedFT). For unweighted models, we used the DAG with the highest AIC score out of 200 unweighted DAGs.

2. *LASSO Multivariate Regularized Regression (RR)*

Each data table was used to fit a multivariate LASSO regression model (sklearn.linear_model.LASSO). For each data table, one of the 0 msec time slice columns served as the predictor variable, with all other columns serving as independent variables (six models corresponding to six 0 ms time slice columns). An edge was considered present if there was a non-zero coefficient from a 2.4 or 1.2 ms time slice variable (edges within the 0 ms time slice were not considered, analogous to DBN analyses). The regularization coefficient was fine-tuned based on the F-score of the resulting model using the *d*=0 data table. We tested a range of regularization coefficients from 0.005 to 0.05 (linearly spaced by 0.005) and found the F-score to peak at α= 0.02.

688 *3. F-Score Calculation and Model Comparisons*

689 To measure how well they conformed to ground truth connectivity, an F-score was calculated for
690 models (MTwDBN, weightedFT, unweighted, RR) fitted to a given subsampled dataset. We use
691 a framework[80] where edges are considered regardless of the time lag in which they appear. The
692 F-Score is defined as: $F = \frac{2RP}{R+P}$, with $R = \frac{C}{C+M}$, $P = \frac{C}{C+I}$, and $C = \#$ connections inferred by the
693 model which are in the ground truth, $M = \#$ connections in ground truth not inferred by DBN, $I = \#$
694 connections inferred by the model which are not in the ground truth. R and P refer to Recall and
695 Precision. In addition to the eight edges indicated in Fig 2e, self-edges were regarded as part of
696 the ground-truth to indicate firing refractory periods. F-Scores of $d \in \{20, 40, 60, 80\}$ were
697 compared by a mixed-model ANOVA with model (MTwDBN, weightedFT, unweighted, RR) as
698 within-subjects variable and *d* (20,40,60,80) as between-subjects variable (RStudio, ezANOVA,
699 revealing significant main effects ($p<0.001$) of model and model x drop interactions after sphericity
700 correction (Greenhouse-Geisser). Models were further compared for each *d* separately using
701 post-hoc Tukey test with Bonferroni correction. MTwDBN is considered to outperform alternative
702 models if it has a higher F-score for all *d* and all $p<0.05$.

703

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

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

710

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

712 To validate whether the edge weights contain additional information about the population
713 dependencies above unweighted edge weights, the binned-and-sliced data table from the
714 synthetic laminar network (*d = 0,* no neurons dropped) was divided into two data tables row-wise,
715 $data_{train}$ (95% of data) and $data_{test}$ (5% of data). $Data_{train}$ was bootstrapped 10 times to mimic multi-
716 session data and separately passed through the MTwDBN pipeline ($B = 50$) to generate 10 x 100
717 weighted DAGs. The set of edges in the weighted graph that survived time-shuffling are denoted
718 E_{set}. Subsampled *unweighted* DAGs containing *n* edges (6 or 13) were obtained by sampling
719 without replacement from E_{set} by either sampling uniformly (UW, unweighted) or sampling using
720 the DAG edge weights as sampling weights (W, weighted). Models UW and W were used for
721 prediction by fitting parameters to 12,000 samples from $data_{test}$ by maximum likelihood
722 estimation[86]. From the remaining samples in $data_{test}$ not used for parameter fitting, 4,000 samples
723 were used for validation. For each sample in the validation set, one of the *effect* variables was
724 chosen at random and both UW and W were used to predict the value of this *effect* variable from
725 the *cause* variables. The prediction accuracies of the two models were compared to the true effect
726 variable value using the M-score[87] . This was repeated for 100 variations of UW and W (created
727 by varying the random seed used to sample edges), and the resulting 100 M-scores were
728 compared using a two-tailed paired t-test for both n=6 and n=13. The W model is considered to
729 outperform UW in prediction if $M_W > M_{UW}$ and $p<0.05$.
730
731 *Attention Data*
732 *a. Behavioral task*
733 Well-isolated single units were recorded from area V4 of two rhesus macaques during an
734 attention-demanding orientation change detection task. The task design and the experimental
735 procedures are described in detail in a previous study[30,88]. While the monkey maintained fixation,
736 two oriented Gabor stimuli were flashed on for 200 ms and off for variable intervals (randomly

737 chosen between 200 and 400 ms). The contrast of the stimulus was randomly chosen from a
738 uniform distribution of 6 contrasts (c = [10%, 18%, 26%, 34%, 42%, and 50%]). One of the stimuli
739 was located at the receptive field overlap region (*Attend In*) and the other at an equally eccentric
740 location across the vertical meridian (*Attend Away*). At the beginning of a block of trials, we
741 presented instruction trials where the monkey was spatially cued to the covertly attend to one of
742 two stimulus locations. One of the two stimuli changed in orientation at an unpredictable time
743 (minimum 1s, maximum 5s, mean 3s). The monkey was rewarded for making a saccade to the
744 location of orientation change. 95% of the orientation changes occur at the cued location, and 5%
745 occur at the uncued location (foil trials). We observed impaired performance and slower reaction
746 times for the foil trials, suggesting that the monkey was indeed using the spatial cue to perform
747 the task. The difficulty of the task was controlled by changing the degree of orientation change
748 (randomly chosen from the following: 1°, 2°, 3°, 4°, 6°, 8°, 10°, and 12°). If no change occurred
749 before 5 s, the monkey was rewarded for holding fixation (catch trial, 13% of trials).
750
751 *b. Electrophysiological recording*
752 While the monkey was performing the attention task (Fig 3a), we used artificial dura chambers to
753 facilitate the insertion of 16-channel linear array electrodes ("laminar probes", Plexon, Plexon V-
754 probe) into cortical sites near the center of the prelunate gyrus. Neuronal signals were recorded,
755 filtered, and stored using the Multichannel Acquisition Processor system (Plexon). Neuronal
756 signals were classified as either isolated single units or multiunit clusters by the Plexon Offline
757 Sorter program.
758 *Laminar boundaries*: For the data collected from linear array electrodes, we used current source
759 density analysis to identify the superficial (Layers 1-3), input (Layer 4), and deep (Layers 5 and
760 6) layers of the cortex based on the second derivative of the flash-triggered LFPs[30,89]. The
761 resulting time-varying traces of current across the depth of the cortex can be visualized as CSD

762 maps (Fig 3b). Red regions depict current sinks in the corresponding region of the cortical
763 laminae, while blue regions depict current sources. The input layer was identified as the first
764 current sink followed by a reversal to current source. The superficial and deep layers had the
765 opposite sink-source pattern i.e. source followed by sink.

766 *SU* classification: Cell bodies of single units with bi-phasic action potential waveforms were
767 assigned to the same layer in which the electrode channel was situated during recordings. Units
768 that had tri-phasic waveforms or other shapes were excluded from analyses. Units with peak-to-
769 trough duration greater than 225μs were classified as broad-spiking putative excitatory neurons;
770 units with peak-to-trough duration less than 225μs were classified as narrow-spiking putative
771 inhibitory neurons (Fig 3c). Extracellular data were collected over 32 sessions (23 sessions in
772 monkey A, 9 in monkey C) yielding 337 single units in total. Unit yield per session was
773 considerably higher in monkey C than monkey A, resulting in a roughly equal contribution of both
774 monkeys toward the population data.

775

776 *c. Data Selection*

777 All analyses in this study were performed on spiking data during an interval of 60-260 m after
778 stimulus onset excluding orientation changes. Only single units whose spike waveforms were
779 successfully classified as broad or narrow and for whom the layer identity could be successfully
780 discerned were used in the analysis. There were 29 sessions which had one or more such units
781 recorded. For layer-wise analyses, only sessions with at least one unit from each layer were
782 included (18 sessions). For broad- and narrow-spiking layer-wise (layer-class) analyses, only
783 sessions with at least one unit from two or more populations were included (27 sessions). Pairwise
784 and network-based dependency analysis were performed on each session separately.

785

786 *d. Analysis: Attention Conditions*

Data from *Attend In* and *Attend Away* trials were analyzed independently. For each attention condition, only data from trials where the animal successfully detected the orientation change or from catch trials where the animal maintained fixation were used.

e. Analysis: Behavioral Performance

We fit the behavioral data with a logistic function and defined the threshold condition as the orientation change that was closest to the 50% threshold of the fitted psychometric function for that session. This subset of trials from within the attend-in condition in which the animal was equally likely to correctly detect (*Hit*) or fail to detect (*Miss*) the orientation change was used for our analysis. Data from *Hit* and *Miss* trials were analyzed independently.

f. Preprocessing for MTwDBN analysis

Single units were grouped according to neocortical layer (superficial/input/deep) for layer-wise (3 populations) analyses and additionally by spike waveform (narrow/broad) for layer+class (3x2 populations) analyses. The “multi-unit” spiking activity of these populations was used for the analysis. The data from each stimulus presentation (60 - 260 ms after stimulus onset) were discretized into 15 ms bins and 6 lags to give 7 time slices (-90, -75, -60, -45, -30, -15, 0 ms). The spiking activity of each population in each bin was discretized to 1 or 0 to denote if there were spikes or not. The data from all stimulus presentations in a session/condition combination were 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,

834 5000 samples were bootstrapped from the binned-and-sliced data tables (see *Data*
835 *Preprocessing*) for each condition separately. To keep analyses consistent across *Attend-*
836 *In/Away* and *Hit/Miss,* the size of the bootstraps was equal to the maximum number of rows of
837 the two conditions to be compared. Logistic regression was performed separately from each
838 *cause* variable to an *effect* variable using the statsmodel python package, BGFS solver[76]. The
839 absolute value of the β_1 coefficient was treated as the total dependency weight to mirror the
840 unsigned dependencies discovered using DBNs. This resulted in 5000 pairs of β_1 across the
841 conditions to be compared. For each pair, attention modulation indices (AMI) and hit modulation
842 indices (HMI) were calculated as:

843 $$AMI = \frac{|\beta_1^{\text{in}}| - |\beta_1^{\text{away}}|}{|\beta_1^{\text{in}}| + |\beta_1^{\text{away}}|}$$

844

845 $$HMI = \frac{|\beta_1^{\text{hit}}| - |\beta_1^{\text{miss}}|}{|\beta_1^{\text{hit}}| + |\beta_1^{\text{miss}}|}$$

846

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

849

850 *i. Calculating Confidence Intervals of Modulation Indices*

851 Modulation indices (either attention or hit) were grouped according to time lag and additionally
852 whether they were between layers (all layers or input ⇔ superficial only) or within layers (all
853 layers, input only, superficial only). For layer-class analyses, they could be additionally grouped
854 according to broad or narrow waveforms. Due to the large sample sizes generated from
855 bootstrapping, hypothesis testing discovers all modulations to be significantly different from zero
856 ($p = 0$). Therefore, an estimation statistics approach was used to estimate confidence intervals of

857 modulation indices. The mean modulation index and bias-corrected and accelerated bootstrap
858 95% confidence intervals were calculated in python (scipy.stats.bootstrap, modified to allow for
859 setting size of bootstrap). To estimate confidence intervals more conservatively given the large
860 sample sizes, the number of resamples and the size of each bootstrap were each set to 5000.
861
862 *j. Generalized phase estimation*
863 To determine the temporal relationship between discrete temporal events and the ongoing local
864 field potential (LFP) signal, we adopted the analytical signal approach for non-stationary wideband
865 signals [90,91]. We first bandpass filtered the LFP from 5-40Hz. This wideband signal captures the
866 dominant fluctuations of the raw LFP signal and at the same time excludes slow global changes
867 and higher frequency signals that could potentially contaminate the LFP. We then calculated the
868 instantaneous phase of this wideband LFP signal (the "generalized phase", GP) and studied the
869 relationship of GP for discrete temporal events of interest (spike times, target stimulus onset time)
870 across experimental conditions. We added a 75ms offset to the target stimulus onset time to
871 account for the average response latency of input layer units in our dataset[92].
872
873 **REFERENCES**

874 1 Ruff, D. A. & Cohen, M. R. Attention Increases Spike Count Correlations between Visual
875 Cortical Areas. *The Journal of neuroscience : the official journal of the Society for*
876 *Neuroscience* **36**, 7523-7534, doi:10.1523/JNEUROSCI.0610-16.2016 (2016).
877 2 Briggs, F., Mangun, G. R. & Usrey, W. M. Attention enhances synaptic efficacy and the
878 signal-to-noise ratio in neural circuits. *Nature* **499**, 476-480, doi:10.1038/nature12276
879 (2013).
880 3 Womelsdorf, T. & Fries, P. The role of neuronal synchronization in selective attention.
881 *Current opinion in neurobiology* **17**, 154-160, doi:10.1016/j.conb.2007.02.002 (2007).
882 4 Bosman, C. A. *et al.* Attentional stimulus selection through selective synchronization
883 between monkey visual areas. *Neuron* **75**, 875-888, doi:10.1016/j.neuron.2012.06.037
884 (2012).

5 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).

6 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).

7 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).

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

9 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).

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

11 Wibral, M., Priesemann, V., Kay, J. W., Lizier, J. T. & Phillips, W. A. Partial information decomposition as a unified approach to the specification of neural goal functions. *Brain Cogn* **112**, 25-38, doi:10.1016/j.bandc.2015.09.004 (2017).

12 Williams, P. L. & Beer, R. D. Nonnegative decomposition of multivariate information. *arXiv preprint arXiv:1004.2515* (2010).

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

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

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

16 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).

17 Knudsen, E. I. in *Annual review of neuroscience* Vol. 30 57-78 (2007).

18 Maunsell, J. & Treue, S. Feature-based attention in visual cortex. *Trends in neurosciences* **29**, 317-322 (2006).

19 Felleman, D. J. & Van Essen, D. C. Distributed hierarchical processing in the primate cerebral cortex. *Cerebral cortex (New York, NY : 1991)* **1**, 1-47 (1991).

20 Freeman, J. & Simoncelli, E. P. Metamers of the ventral stream. *Nat Neurosci* **14**, 1195-1201, doi:10.1038/nn.2889 (2011).

21 Bisley, J. W. & Goldberg, M. E. Neuronal activity in the lateral intraparietal area and spatial attention. *Science (New York, NY)* **299**, 81-86, doi:10.1126/science.1077395 (2003).

22 Treue, S. & Maunsell, J. H. Attentional modulation of visual motion processing in cortical areas MT and MST. *Nature* **382**, 539-541, doi:10.1038/382539a0 (1996).

23 Moran, J. & Desimone, R. Selective attention gates visual processing in the extrastriate cortex. *Science (New York, NY)* **229**, 782-784 (1985).

24 Motter, B. C. Focal attention produces spatially selective processing in visual cortical areas V1, V2, and V4 in the presence of competing stimuli. *Journal of neurophysiology* **70**, 909-919 (1993).

25 Ghose, G. M. & Maunsell, J. H. R. Spatial summation can explain the attentional modulation of neuronal responses to multiple stimuli in area V4. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **28**, 5115-5126, doi:10.1523/JNEUROSCI.0138-08.2008 (2008).
26 Reynolds, J., Chelazzi, L. & Desimone, R. Competitive mechanisms subserve attention in macaque areas V2 and V4. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **19**, 1736-1753 (1999).
27 Treue, S. & Martínez Trujillo, J. C. Feature-based attention influences motion processing gain in macaque visual cortex. *Nature* **399**, 575-579, doi:10.1038/21176 (1999).
28 Williford, T. & Maunsell, J. H. Effects of spatial attention on contrast response functions in macaque area V4. *J Neurophysiol* **96**, 40-54, doi:10.1152/jn.01207.2005 (2006).
29 Hansen, B. J., Chelaru, M. I. & Dragoi, V. Correlated variability in laminar cortical circuits. *Neuron* **76**, 590-602, doi:10.1016/j.neuron.2012.08.029 (2012).
30 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).
31 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).
32 Smith, M. A., Jia, X., Zandvakili, A. & Kohn, A. Laminar dependence of neuronal correlations in visual cortex. *Journal of neurophysiology* **109**, 940-947, doi:10.1152/jn.00846.2012 (2013).
33 Denagamage, S., Morton, M. P., Reynolds, J. H., Jadi, M. P. & Nandy, A. S. Laminar Mechanisms of Saccadic Suppression in Primate Visual Cortex. *bioRxiv* (2021).
34 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).
35 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).
36 D'Souza, R. D. & Burkhalter, A. in *Front Neuroanat* Vol. 11 71 (2017).
37 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).
38 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).
39 Douglas, R. J. & Martin, K. A. C. Mapping the matrix: the ways of neocortex. *Neuron* **56**, 226-238, doi:10.1016/j.neuron.2007.10.017 (2007).
40 Beer, P. L. W. a. R. D. Nonnegative Decomposition of Multivariate Information. *preprint arXiv:1301.6695* (2010).
41 Barrett, A. B. Exploration of synergistic and redundant information sharing in static and dynamical Gaussian systems. *Physical Review E* **91**, 052802 (2015).
42 Tax, T., Mediano, P. A. & Shanahan, M. The partial information decomposition of generative neural network models. *Entropy* **19**, 474 (2017).

43 Tibshirani, R. The lasso method for variable selection in the Cox model. *Stat Med* **16**, 385-395, doi:10.1002/(sici)1097-0258(19970228)16:4<385::aid-sim380>3.0.co;2-3 (1997).
44 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).
45 Koller, D. & Friedman, N. *Probabilistic graphical models: principles and techniques*. (MIT press, 2009).
46 Murphy, K. P. & Russell, S. Dynamic bayesian networks: representation, inference and learning. (2002).
47 Smith, V. A., Jarvis, E. D. & Hartemink, A. J. Evaluating functional network inference using simulations of complex biological systems. *Bioinformatics* **18**, S216-S224 (2002).
48 Zhang, L., Samaras, D., Alia-Klein, N., Volkow, N. & Goldstein, R. in *Advances in neural information processing systems.* 1593-1600.
49 Rajapakse, J. C. & Zhou, J. Learning effective brain connectivity with dynamic Bayesian networks. *Neuroimage* **37**, 749-760 (2007).
50 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).
51 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).
52 Friedman, N., Goldszmidt, M. & Wyner, A. Data analysis with Bayesian networks: A bootstrap approach. *arXiv preprint arXiv:1301.6695* (2013).
53 Levina, A. & Priesemann, V. Subsampling scaling. *Nat Commun* **8**, 15140, doi:10.1038/ncomms15140 (2017).
54 Dalgleish, H. W. *et al.* How many neurons are sufficient for perception of cortical activity? *Elife* **9**, doi:10.7554/eLife.58889 (2020).
55 Priesemann, V., Munk, M. H. & Wibral, M. Subsampling effects in neuronal avalanche distributions recorded in vivo. *BMC Neurosci* **10**, 40, doi:10.1186/1471-2202-10-40 (2009).
56 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).
57 Mitchell, J. F., Sundberg, K. A. & Reynolds, J. H. in *Neuron* Vol. 63 879-888 (2009).
58 Cohen, M. R. & Maunsell, J. H. R. in *Nat. Neurosci.* Vol. 12 1594-1600 (2009).
59 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).
60 McGinley, M. J., David, S. V. & McCormick, D. A. in *Neuron* Vol. 87 179-192 (2015).
61 Landau, A. N. & Fries, P. Attention samples stimuli rhythmically. *Curr Biol* **22**, 1000-1004, doi:10.1016/j.cub.2012.03.054 (2012).
62 Schroeder, C. E., Wilson, D. A., Radman, T., Scharfman, H. & Lakatos, P. Dynamics of Active Sensing and perceptual selection. *Current opinion in neurobiology* **20**, 172-176, doi:10.1016/j.conb.2010.02.010 (2010).

63 Schroeder, C. E. & Lakatos, P. Low-frequency neuronal oscillations as instruments of sensory selection. *Trends in neurosciences* **32**, 9-18, doi:10.1016/j.tins.2008.09.012 (2009).

64 Lakatos, P., Karmos, G., Mehta, A. D., Ulbert, I. & Schroeder, C. E. Entrainment of neuronal oscillations as a mechanism of attentional selection. *Science* **320**, 110-113, doi:10.1126/science.1154735 (2008).

65 Litwin-Kumar, A. & Doiron, B. Slow dynamics and high variability in balanced cortical networks with clustered connections. *Nat Neurosci* **15**, 1498-1505, doi:10.1038/nn.3220 (2012).

66 Yoshimura, Y. & Callaway, E. M. Fine-scale specificity of cortical networks depends on inhibitory cell type and connectivity. *Nat Neurosci* **8**, 1552-1559, doi:10.1038/nn1565 (2005).

67 Lakatos, P. *et al.* Global dynamics of selective attention and its lapses in primary auditory cortex. *Nat Neurosci* **19**, 1707-1717, doi:10.1038/nn.4386 (2016).

68 Haider, B. *et al.* Synaptic and network mechanisms of sparse and reliable visual cortical activity during nonclassical receptive field stimulation. *Neuron* **65**, 107-121, doi:10.1016/j.neuron.2009.12.005 (2010).

69 Lee, T. S. & Mumford, D. Hierarchical Bayesian inference in the visual cortex. *J Opt Soc Am A Opt Image Sci Vis* **20**, 1434-1448, doi:10.1364/josaa.20.001434 (2003).

70 Kersten, D., Mamassian, P. & Yuille, A. Object perception as Bayesian inference. *Annual Reviews in Psychology* **55**, 271-304, doi:10.1146/annurev.psych.55.090902.142005 (2004).

71 Nandy, A., Nassi, J. J., Jadi, M. P. & Reynolds, J. in *Elife* Vol. 8 91 (eLife Sciences Publications Limited, 2019).

72 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).

73 Bertschinger, N., Rauh, J., Olbrich, E., Jost, J. & Ay, N. Quantifying unique information. *Entropy-Switz* **16**, 2161-2183 (2014).

74 Ince, R. A. A. Measuring Multivariate Redundant Information with Pointwise Common Change in Surprisal. *Entropy-Switz* **19**, doi:ARTN 318 10.3390/e19070318 (2017).

75 James, R. G., Ellison, C. J. & Crutchfield, J. P. ``dit``: a Python package for discrete information theory. *Journal of Open Source Software* **3**, 738 (2018).

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

77 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).

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

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

80 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).

81 Ankan, A. a. P., Abinash. in *Proceedings of the 14th Python in Science Conference (SCIPY 2015)* (2015).
82 Akaike, H. in *Selected papers of hirotugu akaike* 199-213 (Springer, 1998).
83 Liu, Z., Malone, B. & Yuan, C. Empirical evaluation of scoring functions for Bayesian network model selection. *BMC Bioinformatics* **13 Suppl 15**, S14, doi:10.1186/1471-2105-13-S15-S14 (2012).
84 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).
85 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).
86 Neapolitan, R. E. *Learning bayesian networks*. Vol. 38 (Pearson Prentice Hall Upper Saddle River, NJ, 2004).
87 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).
88 Nandy, A., Nassi, J. J., Jadi, M. P. & Reynolds, J. Optogenetically induced low-frequency correlations impair perception. *Elife* **8**, e35123 (2019).
89 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).
90 Davis, Z. W., Muller, L., Martinez-Trujillo, J., Sejnowski, T. & Reynolds, J. H. Spontaneous travelling cortical waves gate perception in behaving primates. *Nature* **587**, 432-436, doi:10.1038/s41586-020-2802-y (2020).
91 Muller, L., Reynaud, A., Chavane, F. & Destexhe, A. The stimulus-evoked population response in visual cortex of awake monkey is a propagating wave. *Nature communications* **5**, 3675, doi:10.1038/ncomms4675 (2014).
92 Wang, X., Nandy, A. S. & Jadi, M. P. Laminar compartmentalization of attention modulation in area V4 aligns with the demands of visual processing hierarchy in the cortex. *Scientific reports* **13**, 19558, doi:10.1038/s41598-023-46722-8 (2023).

Figure 1

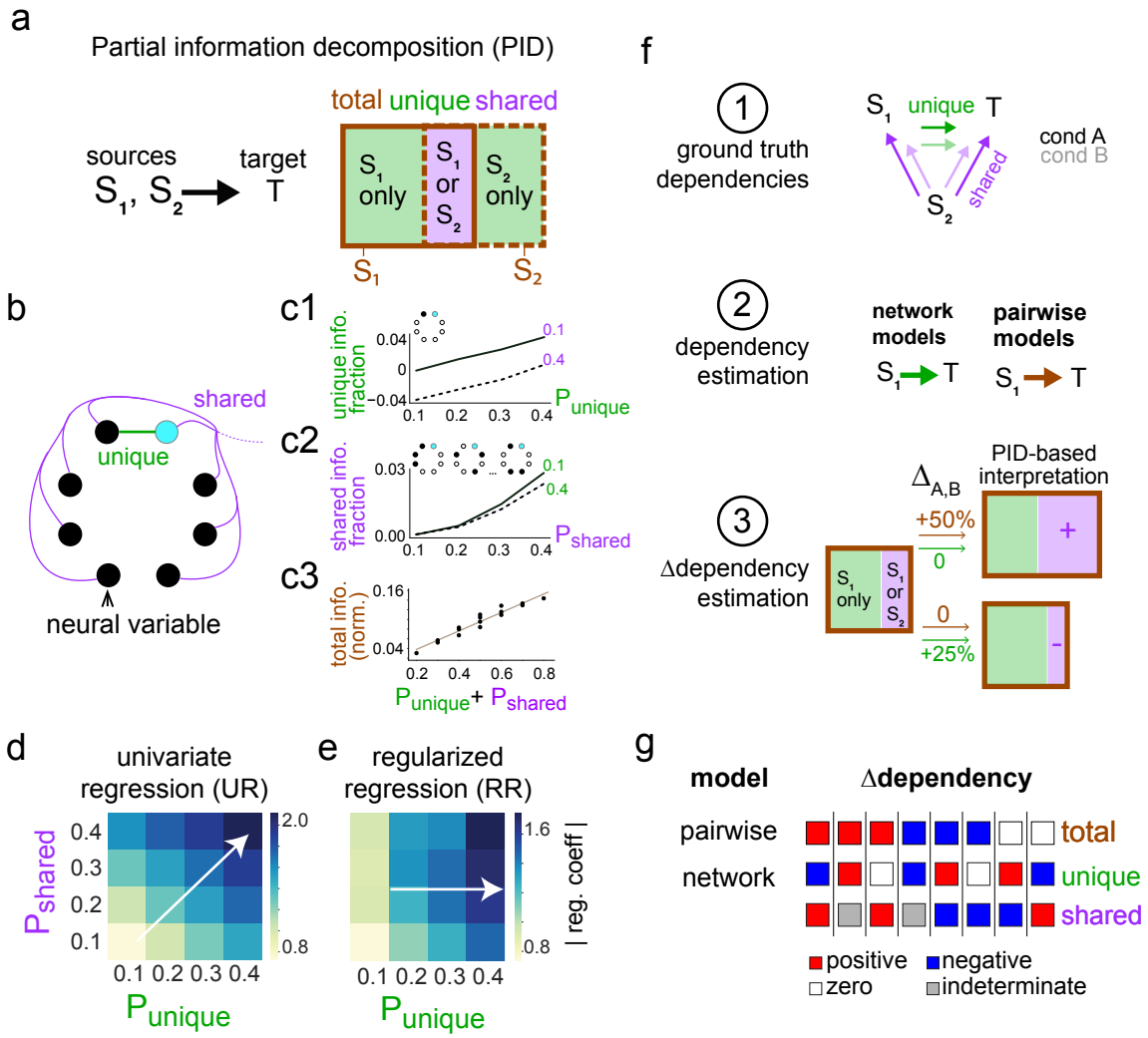

Figure 1 Dependency decomposition in a multi-variate system using pairwise- and network-models.

a Simplified Partial information decomposition (PID) framework[11,12] based definition of types of information that multiple sources can have about a target (see Methods and Supp. Figure FS1).

b 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).

c1, c2 Information fraction (reduction in the proportion of total entropy) as a function of parameters (P_{unique},P_{shared}) that control unique and shared information in the model. Information fraction estimation as a function of P_{shared} (**c2**) was performed using a subset of variables in the simulated network for computational efficiency, (see SI Methods).

c3 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).

d Coefficients of a pairwise model (univariate logistic regression (UR)) as a function of P_{unique} and P_{shared}. White arrow provides a visual guide for direction of highest change in coefficients.

e Coefficients of a network model (LASSO multivariate regularized regression (RR)) as a function of P_{unique} and P_{shared}. White arrow provides a visual guide for direction of highest change in coefficients.

f Application of pairwise and network-based statistical models for approximate information decomposition in an example multivariate system. It illustrates interpretation of the modulation of these dependencies using the PID framework.

g 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) based on the PID framework. Shaded blocks indicate indeterminate modulation direction of shared information in the network.

Figure 2

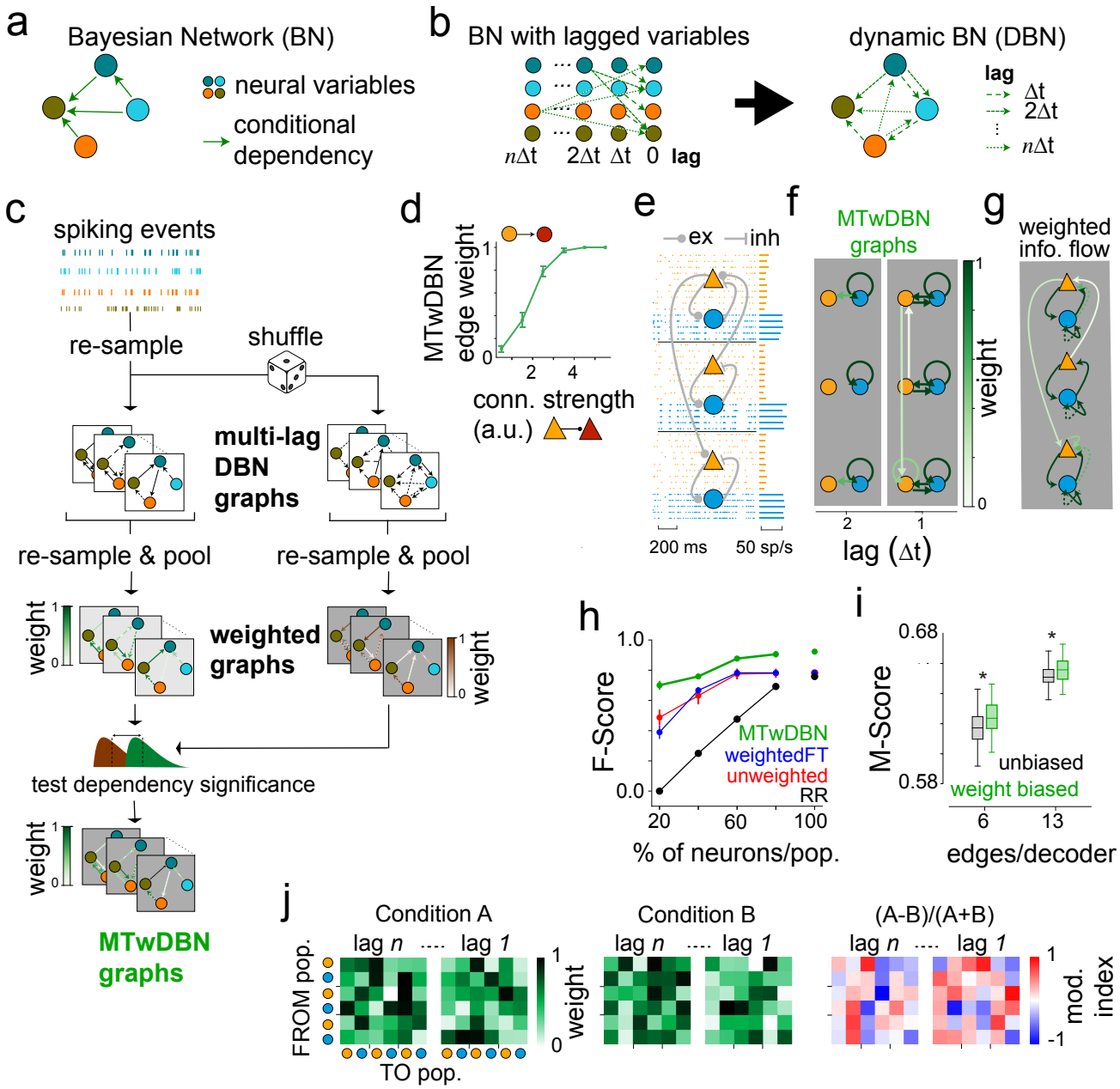

Figure 2 Multi-timescale & weighted Dynamic Bayesian Networks (DBN) based estimation of unique dependencies in a sparsely sampled recurrent neuronal network.
a Bayesian Networks for graph representation of dependencies in a multivariate system. **b** Dynamic Bayesian Networks (DBN) for graph representation of dependencies in multivariate time series data. **c** Analysis flow for multi-timescale weighted DBN (MTwDBN) graphical model fitting. **d** Edge weight of MTwDBN graph as a function of connection weight in a 2-population simulated network using the pipeline in **a**. **e** Spiking activity of 6 subpopulations in a simulated laminar network. Connectivity is visualized in the overlaid schematic. **f** Directed dependencies (edge in the graph) in the simulated network in **e**, estimated using MTwDBN fitting. **g** Summary graph of dependencies across all timescales from **f**. Solid and dashed lines indicated two different timescales. **h** F-score (harmonic mean of precision and recall of dependency structure) as a function of % of neural population observed. F-score was estimated for shuffle corrected weighted DAGs (MTwDBN, green), weighted DAGs with fixed threshold (weightedFT, blue), unweighted DAGs (red), or LASSO regression, an example of regularized regression (RR) models (black). Each point raepresents the average of five separate runs, except 100 % (single run). Error bars indicate standard deviation, some error bars are smaller than symbol size. **i** DBN decoder accuracy with different sizes of MTwDBN DAGs. Decoders were trained to predict population activities using a subsample of shuffle corrected edges (see SI Methods). Graph edges for the decoder were sampled from the learned structure either in an unbiased fashion (black) or biased with the edge weights (green). **j** Schema for estimating modulation of unique dependencies in a network of neural populations, using MTwDBN.

Figure 3

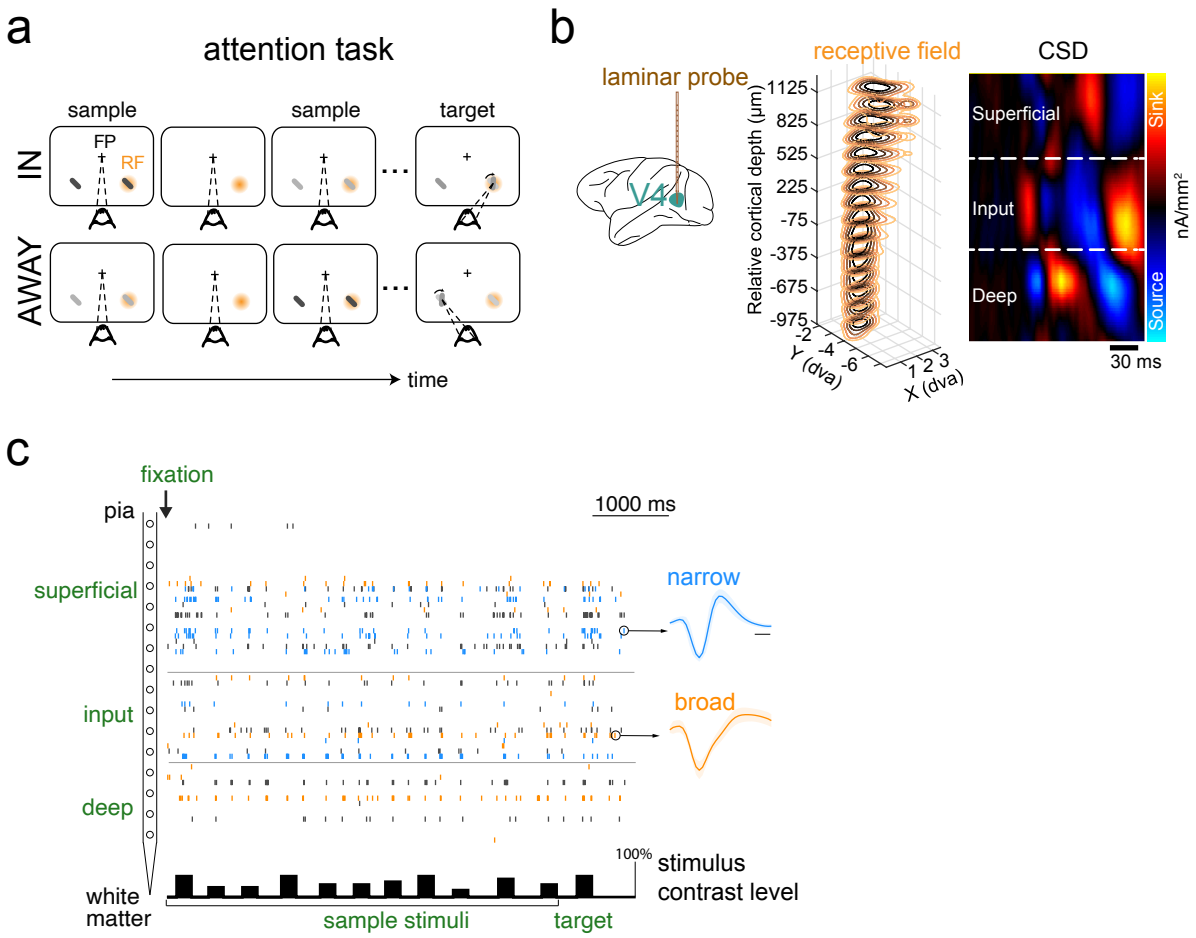

Figure 3 Laminar recordings in area V4

a Experimental protocol: Paired Gabor stimuli with varying 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. Stacked contour plot showing spatial receptive fields (RFs) along the laminar probe from an example session. Alignment of RFs indicates perpendicular penetration down a cortical column. Zero depth represents the center of the input layer as estimated with current source density (CSD) analysis. 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 either single unit (blue, orange) or multi-units (black). Spike waveforms for an example narrow (blue) and broad (orange) single unit are shown.The bars at the bottom depict stimulus presentation epochs, with height indicating relative stimulus contrast.

Figure 4

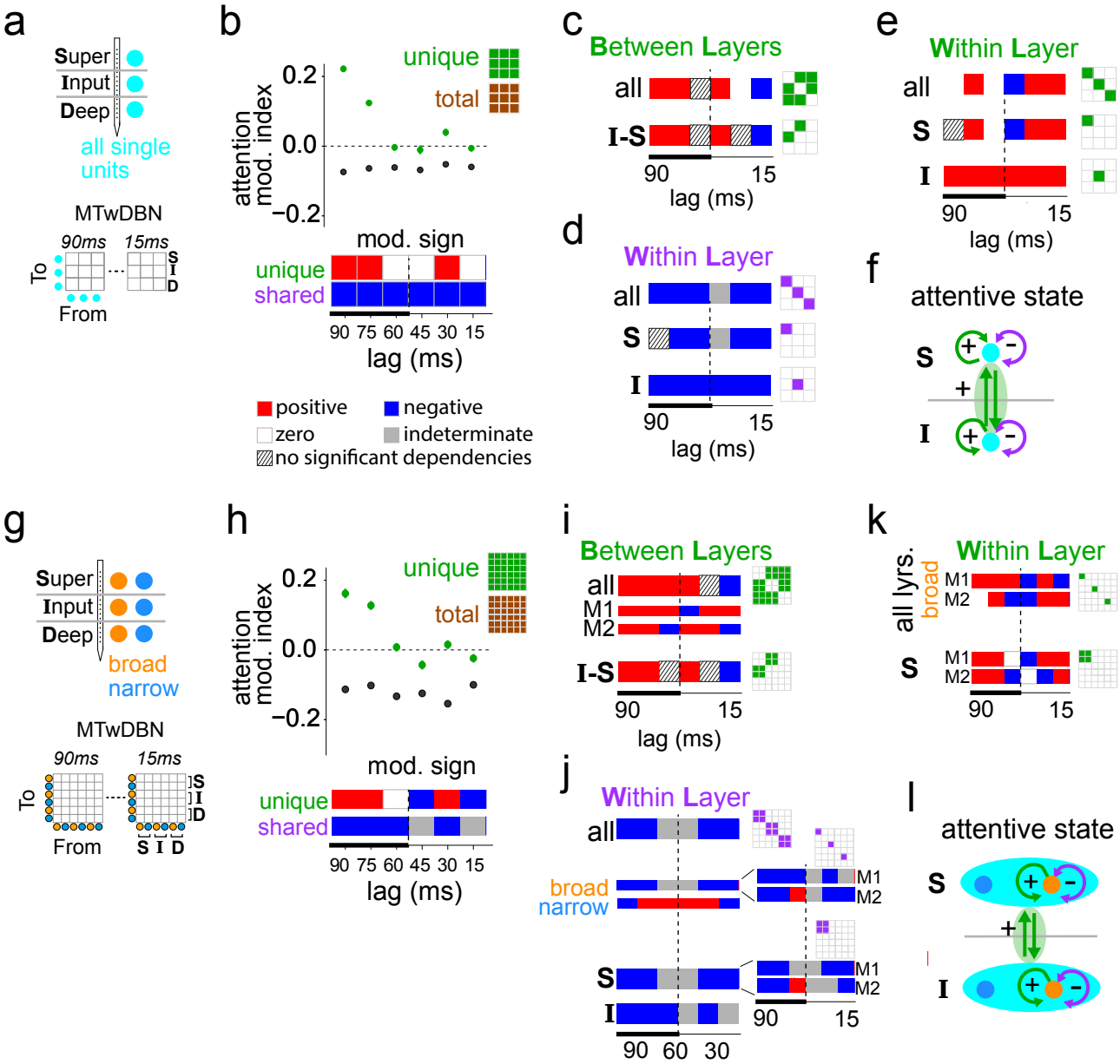

Figure 4 Modulation of dependencies in a V4 laminar network across attention conditions. **a** Neural populations used for fitting laminar MTwDBN. Current source density analysis identified different layers (superficial, input, deep), and isolated single units were assigned to one of these layers (see Methods). **b** Top: Average MTwDBN-based modulation (green) of all unique dependencies between the laminar populations. Modulation of the same dependencies as estimated by logistic regression (brown). Error bars indicate 95% confidence intervals. Bottom: Visualization of modulation sign of unique dependencies at different lags. Combining the modulation sign of unique and total dependencies (see Top), PID framework-based estimated modulation sign of shared dependencies (using schema in Fig 1g) is also shown for different lags. Thicker line along the time axis indicates the timescales of attentional modulation in prior studies38,39. **c** Sign of average modulation of unique dependencies between layers (bi-directionally). I: input layer; S: superficial layer. **d** Sign of average modulation of shared dependencies within layers. **e** Sign of average modulation of unique dependencies within layers. **f** Summary of dependency modulation pattern. **g** Neural populations used for fitting laminar MTwDBN. Isolated single units were classified as broad- and narrow-spiking based on peak-to-trough duration in their average spike shape (see Methods). **h** Top: Average MTwDBN-based modulation (green) of all unique dependencies between the cell-type specific laminar populations. Modulation of the same dependencies as estimated by logistic regression (brown). Bottom: Visualization of modulation sign of unique dependencies and PID framework-based estimated modulation sign of shared dependencies is also shown for different lags. **i** Sign of average modulation of unique dependencies between layers. **j** Sign of average modulation of shared dependencies within layers for all, broad or narrow populations. M1, M2: subject-wise; broad, narrow: cell-class specific. **k** Sign of average modulation of unique dependencies within layers. **l** Summary of dependency modulation pattern. See Fig S3 for modulation indices in **c-e, i-k.**

Figure 5

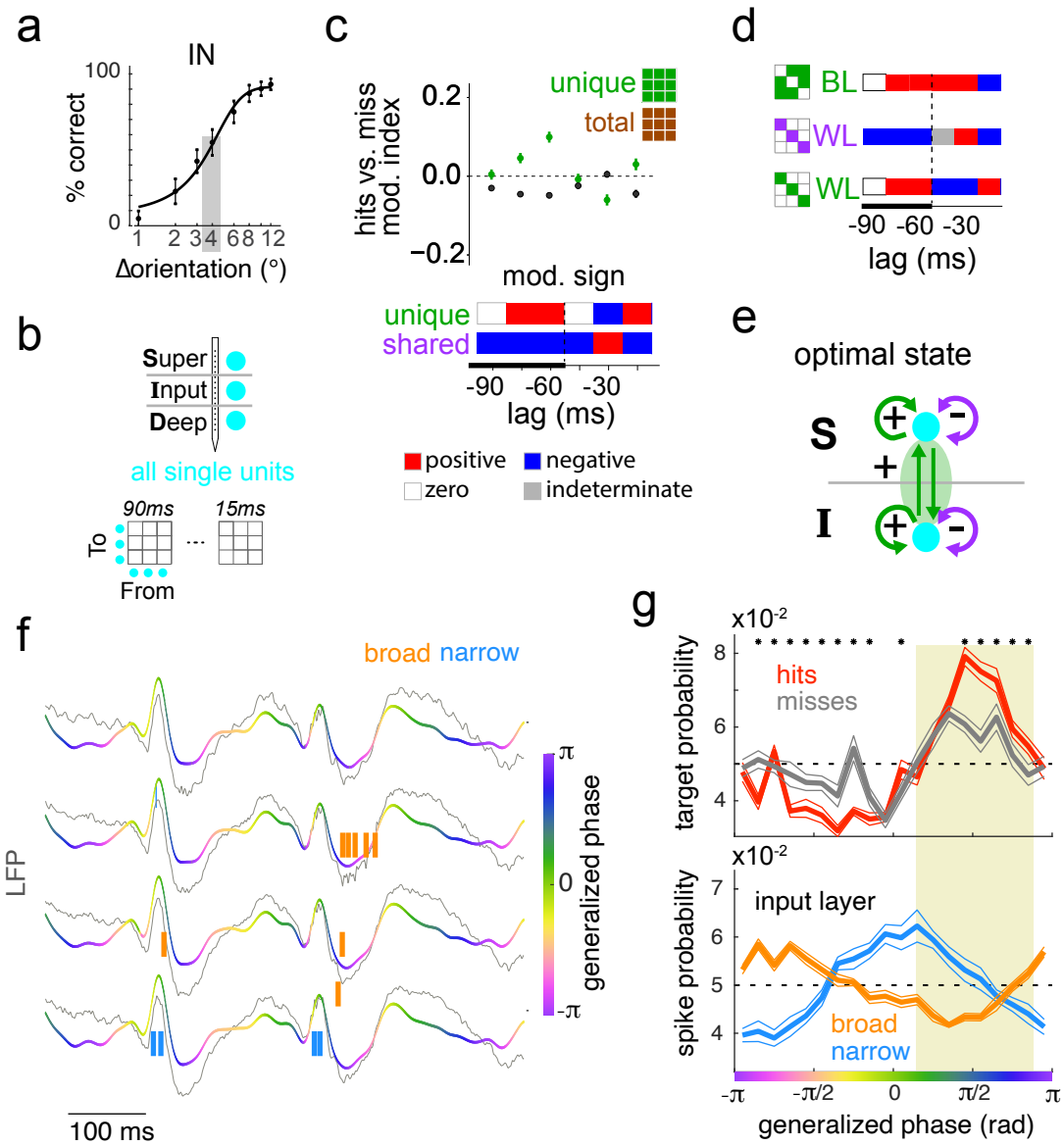

Figure 5 Modulation of dependencies in a V4 laminar network across behavioral outcomes at perceptual threshold. **a** Example session showing performance as a function of task difficulty. Gray box: threshold orientation change at which the animal was equally likely to correctly detect (hit) or fail to 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 (brown) 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. **f** Generalized phase of 5-40 Hz LFP signal in input layer in portion of an example session. Vertical ticks indicate SU spikes (see Methods). **g** *Top*: Target stimulus presentation probability as a function of generalized phase of the LFP (adjusted for cortical delay), separated by HIT and MISS trials. Asterisk (*) indicates phases with significant difference between the two trial types (Ranked sum test, corrected for multiple comparison). *Bottom*: Spike probability in the input layer as a function of generalized phase of the LFP, separately estimated for putative excitatory (broad) and putative inhibitory (narrow) SUs. Error bands indicate Standard Error of Mean. For other layers, see Fig S6.

Figure S1

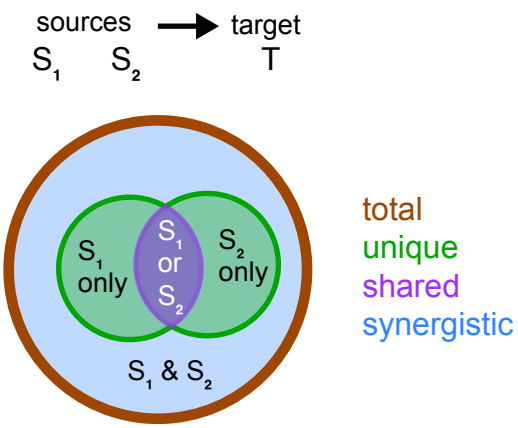

Figure S1 Information decomposition in a multi-variate system
Complete Partial information decomposition (PID) framework[13,14] based definition of types of information that multiple sources can have about a target.

Figure S2

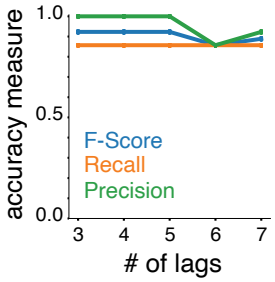

Figure S2 Additional MTwDBN performance metric for synthetic data
Recall, precision and F-score measures for weighted DAGs fit to synthetic data illustrated in Fig 2e as a function of number of lags.

Figure S3

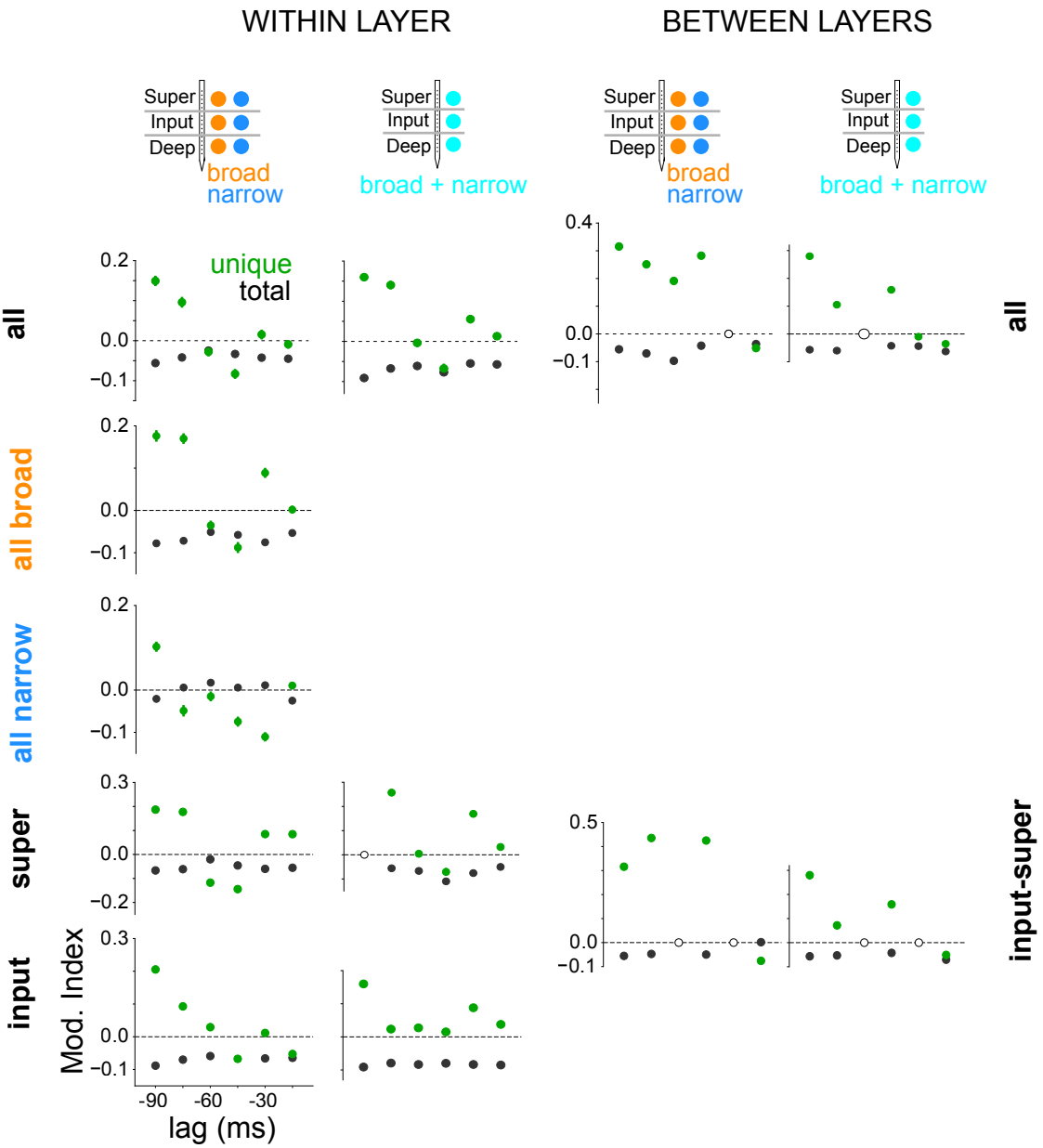

Figure S3 Modulation of dependencies in a V4 laminar network across attention conditions. MTwDBN-based modulation (green) of all unique dependencies between the laminar populations. Modulation of the same dependencies as estimated by logistic regression (black). Open circles indicate no significant edges 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 S4

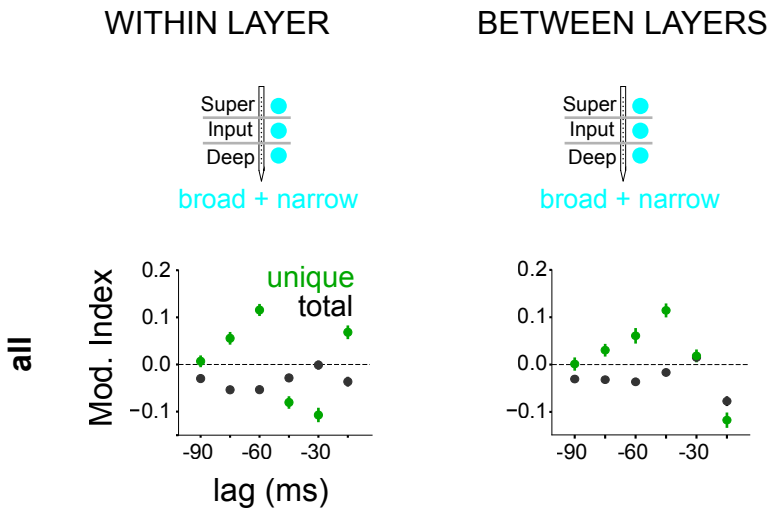

Figure S4 Modulation of dependencies in a V4 laminar network across behavioral outcomes at perceptual threshold. MTwDBN-based modulation (green) of all unique dependencies between the laminar populations. Modulation of the same dependencies as estimated by logistic regression (black). Schematic at the top of every column depicts the populations used for identifying laminar dependency structure within (left) and between (right) cortical layers.

Figure S5

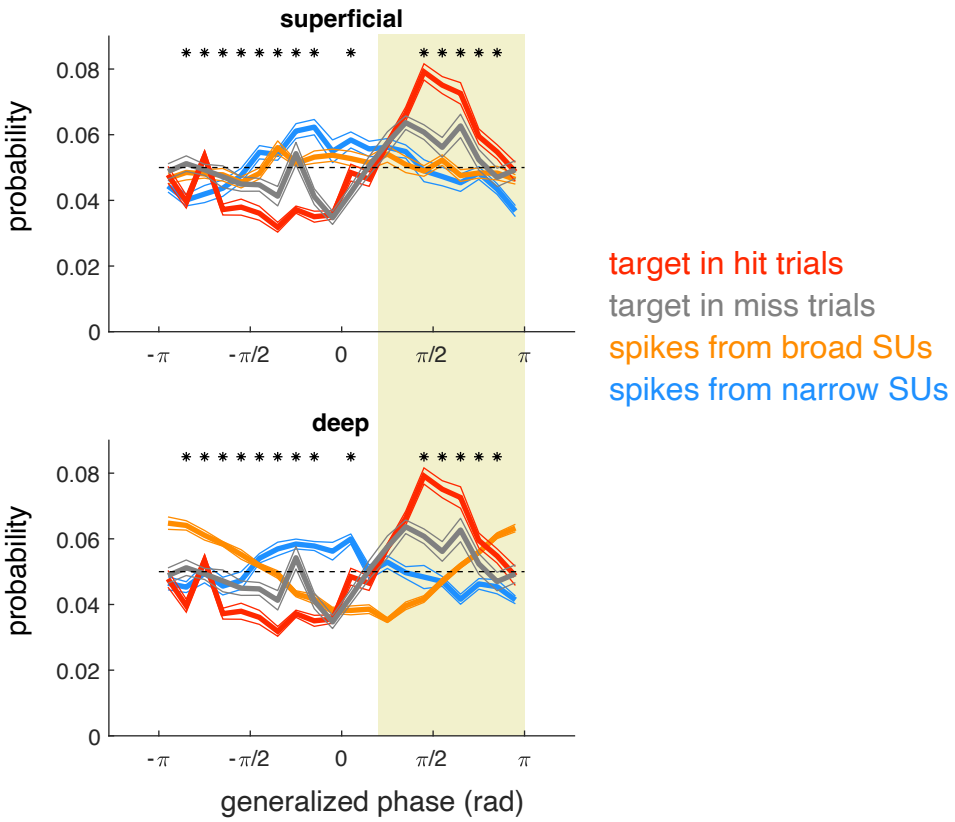

Figure S5 Spike probability in the superficial (top panel) and deep layer (bottom panel) as a function of generalized phase of the LFP, separately estimated for putative excitatory or broad spiking (orange), and putative inhibitory or narrow spiking (blue) SUs. Overlaid is target stimulus presentation probability as a function of generalized phase of the LFP (adjusted for cortical delay), separated by HIT (red) and MISS (grey) trials. Asterisk (*) indicates phases with significant difference between the two trial types (Ranked sum test, corrected for multiple comparison).