Supporting Information for

Information processing dynamics in neural networks of macaque cerebral cortex reflect cognitive state and behavior.

Thomas F. Varley, Olaf Sporns, Stefan Schaffelhofer, Hansjörg Scherberger, Benjamin Dann

Thomas F Varley.
E-mail: tvarley@iu.edu

This PDF file includes:

Supporting Information Text

1. Supporting Figures

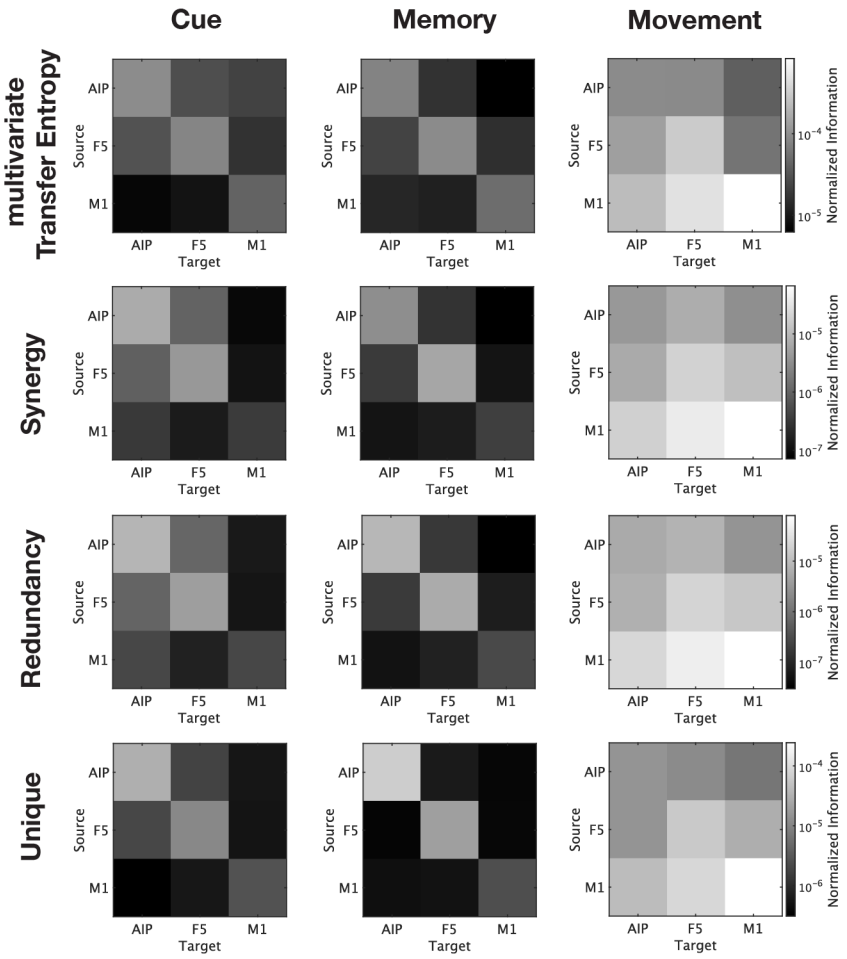

Fig. S1. Behavioral state dependent changes of different types of information processing within and between cortical areas. Displayed as adjacency matrices are the relative strength of the information dynamics: mTE, synergy, redundancy, and unique). Information dynamics within and between the areas (AIP, F5, M1), in the feedforward and feedback direction, are shown for the three behavioral epochs: cue, memory, and movement (averaged across all sliding windows per epoch). Similar to Fig. S1, all displayed information dynamics are averaged across all neural connections within each area or between each area-pair and all recording sessions of all three monkeys. While information processing is largely consistent across all types of measures, large differences are apparent between behavioral epochs. Moreover, during movement execution, information processing is noticeably stronger in the feedback direction than in the feedforward direction.

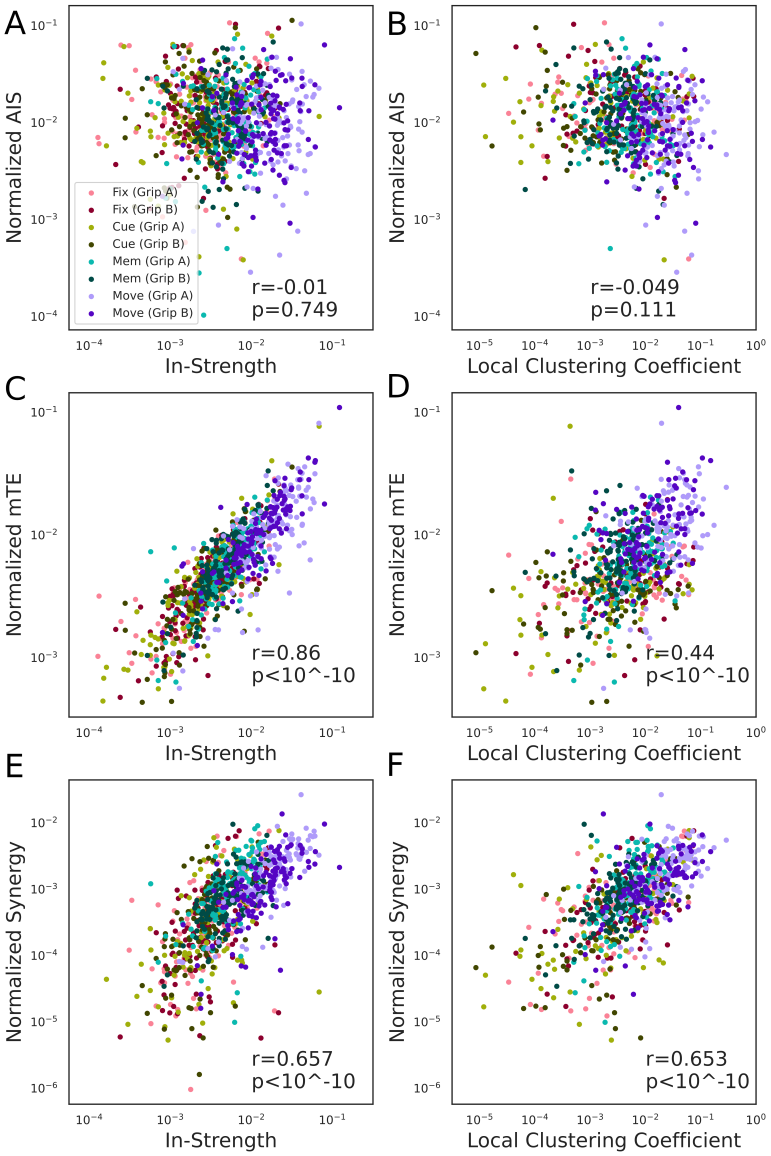

Fig. S2. Relationship of Information dynamics and local network structure. Displayed is how neuron-level information dynamics are related to neuron-level network measures (Spearmanm correlation coefficient). Network measures and information dynamics are averaged across all sliding windows per behavioral period, both grasping conditions and all recording sessions of all monkeys. The color scheme is the same as in Fig. 4 **A-B** We found no significant relationship between active information storage and the in-strength of a neuron or the local clustering coefficient. **C** There was a highly significant correlation between the average transfer entropy of a neuron and its in-strength ($r = 0.869, p < 10^{-10}$), which is unsurprising given that both measures quantify the total information flow into a neuron, albeit in slightly different ways. **D** However, there was a significant correlation between average transfer entropy and local clustering coefficient ($r = 0.44, p < 10^{-10}$), which shows that the local community structure of a neuron reflects the associated intensity of information flow. **E** The triadic synergy was positively correlated with the in-strength of the target neuron ($r = 0.657, p < 10^{-10}$), replicating the findings of (1), as well as with the local clustering coefficient ($r = 0.653, p < 10^{-10}$), suggesting that the extended local neighborhood structure plays an important role for computation.

2. Supplementary Materials & Methods

A. Behavioral Task. Fig. 1 in the main text illustrates the time course of the behavioral task described in detail previously (2, 3). Monkeys were trained to perform a mixed free-choice and instructed delayed grasping task, with the exception of monkey M, which was trained to perform only a instructed delayed grasping task. They were seated in front of a handle that could be grasped in two different ways and visual cues were displayed on a masked monitor that was superimposed on the handle using a beam splitter mirror. Trials started after the monkey placed both hands (for monkey S) or one hand (for monkey Z and M) on the resting positions. For monkey Z and M, the non-task relevant arm was comfortably restrained in an acrylic glass tube during the experiment. Next, a small red disk was displayed centrally on the monitor, which had to be fixated for a variable period of 600 to 1000ms (fixation epoch). After the fixation epoch, cues in the form of larger disks were shown next to the fixation disk for 300ms and the handle was illuminated. In the instructed context, one of two discs was displayed instructing the monkey about the required grip type (power or precision; cue epoch). In the free-choice context both disks were

displayed indicating the monkey to freely choose between the two grip types. After the cue was turned off, the monkey had to remain steady for a variable time of 1100 to 1500ms (memory epoch). The turning off of the central fixation disk instructed the monkey to reach and grasp the target with the required grip type (movement epoch) to receive a liquid reward. Note that to encourage the monkey to perform both grip types during the free-choice context, the reward was iteratively reduced every time the monkey repeatedly chose the same grip type (fraction of power grip selections: 38.5±4.2% SD and 53.3±5.2% SD for S and Z, respectively; Note that no choice ratio is available for monkey M, since he performed only instructed trails). All trials were randomly interleaved and executed in darkness. Note that the behavioral task also contained delayed instructed trials for monkey S and Z, which were not analyzed in this study. To increase the reliability of the information dynamics estimation, instructed and free choice trials for the same grip type were pooled for all further analyses.

B. Electrode Implantation and Data Recordings. Surgical procedures have been described in detail previously (2, 3). Briefly, two floating microelectrode arrays (FMAs; Microprobes for Life Sciences; 32 electrodes; spacing between electrodes: 400μm; length: 1.5 to 7.1 mm monotonically increasing to target grey matter along the sulcus) were implanted per area in the ventral premotor cortex (area F5), in the anterior intraparietal area (AIP), and in monkey M also in the primary motor cortex (M1) for a total of 128 electrodes for monkey S and Z, and 192 electrodes for monkey M. Monkeys S and M were implanted in the left and monkey Z and the right hemispheres (Fig. 1.B. in the main text).

Extracellular signals from the implanted arrays were amplified and digitally stored using a 128-channel recording system for monkey S and Z (Cerebus, Blackrock Microsystems, sampling rate 30 kS/s, 0.6-7500Hz band-pass hardware filter) and a 256 channel Tucker-Davis system for monkey M (TDT RZ2; sampling rate 24.414 kS/s, 0.6–10,000 Hz band-pass hardware filter) while the monkeys performed the delayed grasping task. All data were saved to disk and analyzed using custom MATLAB code (The Mathworks Inc., Natick, MA).

C. Data Preprocessing. For spike detection, the broadband signals were first low-pass filtered with a median filter (window length 3ms) and the result subtracted from the raw signal, corresponding to a nonlinear high-pass filter. The signal was then low-pass filtered (4th order non-causal Butterworth filter, fc: 5000 Hz). To eliminate common noise sources, principal component (PC) artifact cancellation was applied for all electrodes of each array, as described previously (3, 4). To ensure that no true spikes were eliminated from individual channels, PCs with any normalized coefficient greater than 0.36 (conservatively chosen) were retained. Spike waveforms were detected and semi-automatically sorted using a modified version of the offline spike sorter WaveClus (3, 5, 6).

Units were classified as single- or non-single unit based on five criteria:

1. The presence of short (1–2ms) intervals in the inter-spike interval histogram for single units;
2. The similarity of all waveforms of each unit at each time point;
3. The separation of waveform clusters in the projection of the first 13 features (a combination for optimal discriminability of PCs, single values of the wavelet decomposition, and samples of spike waveforms) detected by WaveClus;
4. The presence of well-known waveform shapes characteristics for single units;
5. The shape of the inter-spike interval distribution.

After the semiautomatic sorting process, redetection of the different average waveforms (templates) was done to detect overlaid waveforms (3, 7). To achieve this, filtered signals were convolved with the templates starting with the biggest waveform. Independently for each template, redetection and resorting was run automatically using a 32-dimensional linear discriminate analysis for classification of waveforms. After spike identification, the target template was temporally matched and subtracted from the filtered signal of the corresponding channel to reduce artifacts during the detection of the next template. This procedure allowed us to detect spikes with a temporal overlap up to 0.2ms. Unit isolation was evaluated again, based on the five criteria mentioned above, to determine the final classification of all units into single or non-single units. Stationarity of firing rate was checked for all units and in case it was not stable over the entire recording session (more than 30% change in firing rate between the first 10 min and the last 10 min of recording) the unit was excluded from further analyses ($< 3\%$ of all single units). Note that only well isolated single neurons based on the five criteria above were used for all further analysis.

For the evaluation of the average firing rates of individual neurons per condition and over time, spike events of all well isolated neurons were binned in non-overlapping 1-ms windows resulting in continuous spike count signals sampled with 1 kHz. Next, spiking signals were low pass filtered with a Gaussian window (= 50 ms) and aligned to cue onset and movement onset (cue onset: -450 to 1250 ms, and movement onset: -450 to 150 ms). Data were aligned to cue and movement onset because neural activity was locked to both events and because the variable memory period length differed between trails. The time range of the alignments was chosen in order to have an as small as possible overlap and an as smooth as possible transition between both periods.Finally, average firing rates were calculated by averaging over all trials and alignments per condition and neuron.

D. Sliding Window Analysis. To assess how effective connectivity patterns change with the behavior of the macaque, we made use of a sliding-windows analysis (8). First, spike events of all well isolated neurons were binned in non-overlapping 5-ms bins resulting in a sampling rate of 200Hz. Second, we used 800 ms windows, incremented with 100 ms tied to either the onset of the cue condition (starting with -800ms from the onset of the cue and continuing for 1700 ms), or the onset of movement (starting with -800 ms from the onset of movement and continuing for 600 ms.) The sliding windows were aligned to the cue and movement onset due to the reasons mentioned above.The result was ≈900 individual slices across the 10 different recordings.

For each of these slices, we inferred a single multivariate transfer entropy network using the IDTxl package (9), as well as the local active information storage, and triadic synergies following the pipeline outlined in (1, 10).

We chose the sliding window approach to temporal localization instead of the more fine-grained analysis of local information dynamics (11) to ensure that we could see how the overall network topology changed in time: a local information approach would have improved our temporal resolution but at the cost of providing a static network, since fixation, cue, and movement conditions would need to be aggregated to construct a single joint probability distribution (which the system presumably visits different parts of in different conditions, which the local analysis would reveal). This would have precluded a dynamic network approach, which is a core aspect of this study.

E. Active Information Storage. Active information storage (AIS) is arguably the simplest of the three information dynamics explored here, and quantifies how much information about the next state of the neuron is encoded in the past k states (11). AIS can be thought of as a measure of the "memory" capacity of the neuron: if every subsequent state is chosen at random, then $AIS = 0$, as knowing the past does nothing to reduce our uncertainty about the mediate future, however if the past states of the neuron affects the probability of a particular next state, then the system can be thought of as "remembering" that information and "using" it when deciding what the next state will be (note the scare quotes). A simple example might be the refractory period that follows an action potential: following an action potential, most neurons cannot immediately fire again, as the correct charge needs to be built up. Consequently, an observer attempting to model the dynamics of a single neuron would know that the probability of a spike goes down if the neuron is known to have fired in some previous time-window, and this would be reflected in the AIS measure. For a more involved discussion of AIS in understanding neural dynamics see (12).

The AIS of a single variable X at time t is given as the mutual information between the immediate state of X at time t and the embedded joint-state of the past k time-steps (the embedding here is similar to a Takens embedding (13)):

$$AIS^{-k}(X) = I(X_{t-1}^{-k}; X_t) \quad [1]$$

Intuitively, this can be understood as measuring how much knowledge of the joint past reduces our uncertainty about the immediate future. The free parameter k indicates how many steps into the past we want to consider when calculating the AIS and should be chosen with some care. If k is too small, then we run the risk of missing out on relevant information from more than k steps in the past and we will under-estimate the true AIS. On the other hand, if k is too large, then we run the risk of under-sampling the joint-probability space and over-estimating the true AIS.

In this work, we set *max k=3*, which given 5 ms bins works out to a considered history of 15 ms (based on prior work reported in (14, 15)), and then used the IDTxl package (9) to infer an optimal non-uniform embedding. The IDTxl package has built-in algorithms for selecting an optimal embedding based on hierarchical significance testing and min/max statistical tests (for detail see (16) and the IDTxl documentation. Briefly, the significance of the AIS was computed by comparing the empirical value to a null distribution of surrogate time series, where were generated by shuffling the empirical time series in 25 ms windows. This localized shuffling randomizes short-term correlations between neurons, while preserving long-term variability in firing rates and helps correct for nonstationarities in the data. We used 2000 shuffled nulls for the min/max statistics, and 5000 shuffled nulls each for the omnibus and maximum sequence testing. Following (10, 17), the AIS of a neuron was normalized by dividing it by the total Shannon entropy of the neuron. This provides a simple correction for variable firing rates, removing the confound that higher-entropy neurons have a greater capacity for any information dynamic.

F. Multivariate Transfer Entropy Networks. While the AIS is restricted to understanding the information dynamics of a single neuron (how it is "remembered" through time), the transfer entropy (TE) describes how information "flows" from one source neuron (or a set of source neurons) to a single target neuron (11, 18, 19) For a source X and a target Y, $TE(X \to Y)$ is given by the mutual information between X's past and Y's immediate next state, conditioned on Y's own past.

$$TE^{k,l}(X \to Y) = I(X_{-k:t-1}; Y_t | Y_{-l:t-1}) \quad [2]$$

By conditioning on Y's past, we are able to extract the information that X's past provides about Y's future *above and beyond the AIS in Y*. Transfer entropy has become an established tool in modeling neural data (1, 20–23). A common technique is to use transfer entropy to construct an effective connectivity graph meant to model the directed synaptic structure of the underlying biological network, to reveal properties like rich clubs (10, 24), multiplex structures (25), and communities (26).

The "classical" transfer entropy has a significant limitation, however that it is strictly bivariate, and consequently blind to higher-order information dynamics. Consider the case of two source neurons A, B synapsing onto a single single target neuron C, which implements $C_t = XOR(A_{t-1}, B_{t-1})$. Due to the synergistic nature of the logical XOR, bivariate transfer entropy will not find any relationship between A, B and C, since information about C_t is only disclosed by the joint pasts of A and B considered together. To address this, we can use the conditional transfer entropy, which quantifies the information flow from one neuron to another *in the context of other neurons in the system*. Formally:

$$mTE^{k,l,m}(X \to Y|\mathbf{Z}) = I(X_{-k:t-1}; Y_t|Y_{-l:t-1}, \mathbf{Z}_{-m:t-1}) \quad [3]$$

Where $\mathbf{Z}$ can be a multidimensional variable. Note again that each variable can have a different lag, to account for differences in the intrinsic dynamics in each element. In our simple logical XOR example, $TE(A \to C) = 0$ bit, but $TE(A \to C|B) = 1$ bit.

For a would-be network neuroscientist, this presents an optimization problem: for a pair of neurons X and Y, how do we find the smallest set of neurons Z such that the context Z provides discloses *all* the relevant information flow from X to Y. The answer is the multivariate transfer entropy (mTE inference, detailed in (16), which uses greedy optimization and multi-level statistical hypothesis testing to define a "parent" set for every neuron in the network, from which the optimal conditional transfer entropy between every parent in the parent set and the target neuron can be derived. Significance testing was done using the same hierarchical testing described above, with null distributions constructed from shuffled time series in 25ms windows. Work using simulated data has shown that the bivariate and multivariate transfer entropies return networks with strikingly different topologies and that the multivariate transfer entropy networks are reliably closer to the "ground truth" (27, 28). It has been established previously that non-trivial synergistic information dynamics exist in biological neural networks (10, 17, 29), suggesting that bTE, which is known to be blind to synergistic relationships, is missing important aspects of the system's dynamics.

As with the *AIS*, the choice of embedding parameters is important, and to maintain consistency, we again set the maximal k value to 3, and used the IDTxl package (9) to find optimal non-uniform embeddings for each neuron. Due to the large number of neurons recorded, we inferred the parent set of each neuron independently and constructed a putative mTE network post-hoc without correcting for familywise error rates at the whole-network level (every edge was corrected for multiple comparisons). This is an acknowledged limitation of our inference, however the computational requirements for a full, error-corrected inference are beyond our capabilities at present. It may be possible that a subset of edges inferred are false-positives, although we do not anticipate that enough false positives would have passed the otherwise stringent threshold to substantively compromise the overall analysis.

F.1. Density, Rich Club Coefficient & Clustering Coefficient. For a directed network $G = (V, E)$, where V is the vertex set and E is the edge set, the density of the network is given as:

$$\frac{|E|}{|V|^2 - |V|} \quad [4]$$

Which records the ratio of the number of edges that exist in the network to the total number of edges that could exist after self-loops are removed. For processing networks of neurons, such as used in this study, network density reflects the amount of effective connections between neurons involved in processing at a given time and condition.

For a directed, weighted graph, the local clustering coefficient for a node v is given by:

$$c_v = \frac{1}{deg(v)(deg(v) - 1) - 2deg^{-1}(v)} T(v) \quad [5]$$

where $deg(v)$ is the sum of the in- and out-degrees of v, deg^{-1} is the reciprocal of the degree of v, and $T(v)$ counts the number of directed triangles that v participates in. We used the Networkx clustering toolbox (30) for this calculation. The average clustering coefficient for the whole network is given by:

$$C(G) = \frac{1}{|V|} \sum_{v \in V} c_v \quad [6]$$

Which is simply the unweighted average of each local clustering coefficient. Finally, the average clustering coefficient was normalized by the corresponding average of surrogate clustering coefficients. surrogate clustering coefficients were estimated based on 1000 surrogate networks, as described in detail in the next subsection. For processing networks of neurons, such as used in this study, the local clustering coefficient reflects groups of effectively interconnected neurons involved in processing at a given time and condition.

The rich club coefficient quantifies the extent to which nodes of a network with a high number of connections are more strongly connected with each other then expected by chance (31, 32). We calculated a modified version of the original weighted rich club coefficient as described in detail in the following (10). The weighted rich club coefficient was calculated separately for all effective mTV networks (23 time windows x 2 grasping conditions). First, all neurons per networks were sorted by the combined in and out degree to obtain a degree rank r_{deg} per neuron. Second, the weighted rich club coefficient $\Phi_w(r_{deg})$ was consecutively calculated given by:

$$\Phi_w(r_{deg}) = \frac{W_{r_{deg}}}{r_{deg}^2 - r_{deg}} \quad [7]$$

with $W_{r_{deg}}$ denoting the sum of weighted connections of the subset of neurons with a degree rank $> r_{deg}$, and with $r_{deg}^2 - r_{deg}$ denoting the maximum possible number of connections of the same subset. Since neurons with a high degree have an increased probability by chance to be connected to other neurons with a high degree, the weighted rich club coefficient needs to be

normalized by surrogate data based on networks with the same degree and ideally strength distribution. The estimation of the surrogate networks is described in detail below. In a third step, each rich club value was normalized by the corresponding average of 1000 surrogate weighted rich club coefficients. Finally, the normalized weighted rich coefficients were resampled to 1 to 100 using linear interpolation to average across different recording sessions with different numbers of neurons. For processing networks of neurons, such as used in this study, the normalized rich club coefficient shows the degree to which a a central processing center is present at a given time and condition.

F.2. Surrogate Networks. To estimate the normalized cluster coefficient, the normalized rich club coefficient and for statistical purposes we generated 1000 partitions of surrogate networks for each inferred multivariate transfer entropy network (21 time windows × 2 conditions × 10 recording sessions). Surrogate networks were generated by first shuffling network connections and second reassigning connection weights. Networks of neurons estimated from simultaneous extracellular recordings are constrained by multiple anatomical and technical factors. To ensure that the above network analyses and statistics are not biased by these factors, network properties affected by these factors should be preserved in the surrogate networks. Several studies have shown, that the number and strength of neuronal connectivity decreases with spatial distance (3, 33, 34). Additionally, the spatial electrode configuration on each array as well as the configuration of the recording arrays to each other results in a spatial inhomogeneous subsampling of neurons (3, 35). To account for both, we held the number of connections on the same electrode, the same array, the same cortical area, and the different inter-area connections the same and matched the average connection strength per distance category as close as possible. Cortical neurons also have an increased probability to be reciprocally connected to each other (36). For this reason, we also held the ratio of reciprocal to unidirectional connections per distance category the same. Furthermore, for the normalization of the rich club coefficient it is necessary that the number of connections per neuron is preserved because strongly connected neurons have an increased probability of being connected to each other by chance (31, 32). However, the same is true for the connection strength per neuron, which has been ignored in the literature so far. We therefore not only held the number of connections per neuron the same, but also matched the connection strength per neuron as close as possible. To our knowledge the described above surrogate network method is more conservative than any other method used in the field, which emphasizes the validity of the results based on this surrogate method.

F.3. Cluster-Based Significance Testing. For the statistical testing of increased rich club and clustering coefficient over the time course of the task, we performed cluster-based surrogate tests. Statistical testing was performed across recording sessions and separate for the power and precision condition (Fig. 1.A. in the main text) (3, 37). The testing procedure was the same in both cases and the same set of surrogate networks was used as described above. The computation of the cluster-based surrogate tests was performed as following:

1. Test values (rich club coefficient or rich club coefficient) were calculated separately for all surrogate networks of all time windows and recording sessions.

2. Normalize all surrogate test values by dividing the average surrogate test values of the corresponding time window and recording session (normalized rich club coefficient or normalized rich club coefficient).

3. Average all surrogate test values per partition and time window (for the normalized rich club coefficient the interval between the 83rd and 95th percentiles) to obtain one value per network.

4. Calculate the surrogate t-values across recording sessions for all partitions and time windows.

5. Select all surrogate t-values larger than a thresholding criterion and cluster and sum them on the basis of temporal adjacency.

6. Take the largest summed surrogate t-value per partition to construct a distribution of largest summed surrogate t-values.

7. Repeat step 1-5 for the recorded test values.

8. For every summed t-value calculate the proportion of surrogate t-values that are larger than the recorded summed t-values, which corresponds to the p-value.

9. Compare each p-value with a critical alpha-level (0.05 in all cases).

Note that this single comparison replaces the multiple comparisons of the test-values over time.

G. mTE and AIS Distribution Similarity Analysis. Similarity matrices derived from the time-resolved TE and AIS analysis underwent clustering and community detection to extract, in a purely data-driven way, time periods that had a distinct TE similarity structure. To create similarity matrices for the distribution of AIS values across the nodes, we created a matrix where cell ij gives the cosine similarity between the vector of AIS values across all nodes in the i^{th} and j^{th} windows. Cosine similarity is given as:

$$Sim(X,Y) = \frac{\langle X,Y\rangle}{||X|| \times ||Y||} \quad [8]$$

Where $\langle X, Y \rangle$ is the dot product of X and Y, and $||X||$ is the norm of X. The implementation we used was from the Scikit-Learn Pairwise Comparisons package.

The mTE pipeline was the same, although we took the cosine similarity of the flattened adjacency matrices. First, similarity matrices were aggregated across the 10 trial repetitions by taking the mean. This resulted in a single 46x46 element matrix for both mTE and AIS. These matrices were then clustered using a version of modularity maximization (employing the Potts null model) and multi-resolution consensus clustering (38). The resolution parameter, used by the Louvain algorithm, was stepped through a range of 1,000 values covering modular partitions yielding between 2 and 46 modules (the minimal and maximal number of modules possible). The resulting 1,000 partitions were aggregated into a single co-classification matrix which was scaled between followed by subtraction of an analytic null that captures the expected co-classification (mean propensity for each node pair to be grouped in the same community) under random permutation of the module assignments (38). The resulting scaled co-classification matrices were then subjected to consensus clustering with $\tau = 0$. The level of τ corresponds to the (constant) null. The resulting consensus communities correspond to time steps that are clustered together based on the similarity of their mTE/AIS similarities. Variations in the number of samples (between 100 and 10,000) and range of resolution parameter (2-23; 2-10) had no effect on the cluster boundaries.

H. Synergy & Partial Information Decomposition. The final information dynamic is *information modification*, which describes information that is somehow non-trivially altered by the interaction of two or more incoming "information streams" (11). We operationalize this notion with the synergy, which represents information that is, in some sense, "greater than the sum of its parts."

Synergy is defined as all the information about a target Y that can only be learned when observing two sources, X_1 and X_2 jointly. The *partial information decomposition* provides a framework for doing this, decomposing the joint mutual information $I(X_1, X_2; Y)$ into:

$$\begin{aligned} I(X_1, X_2; Y) = &\text{Red}(X_1, X_2; Y) + \text{Syn}(X_1, X_2; Y) + \\ &\text{Unq}(X_1; Y|X_2) + \text{Unq}(X_2; Y|X_1) \end{aligned} \quad [9]$$

Where Red() indicates the information about Y that can be extracted from both X_1 or X_2, Unq() indicates the information that is uniquely disclosed by only X_1 or X_2, and Syn() indicates the information about Y that is disclosed by the joint states of X_1 and X_2 together and no simpler combination of elements.

While PID provides an elegant structure with which to understand the decomposition of information, it does not specify how exactly to calculate any of the desired values. We used the built-in PID solver provided by IDTxl.

H.1. Redundancy / Synergy Bias. It is known that different systems can be variously dominated by redundant or synergistic information dynamics (39, 40). We hypothesized that the fixation/cue states would have a different ratio of synergistic and redundant information dynamics. The idea of an explicitly normalized redundancy/synergy ratio has been proposed in (40), who found that topologically similar systems can nevertheless have strongly different distributions of partial information over the PI lattice. To calculate the redundancy/synergy ratio (RSR), we normalize the relevant PI-atoms by the overall joint mutual information:

$$\bar{S}(X_1, X_2; Y) = \frac{Syn(X_1, X_2; Y)}{I(X_1, X_2; Y)} \quad [10]$$

$$\bar{R}(X_1, X_2; Y) = \frac{Red(X_1, X_2; Y)}{I(X_1, X_2; Y)} \quad [11]$$

The ratio is then:

$$RSR(X_1, X_2; Y) = \log\left(\frac{\bar{R}}{\bar{S}}\right) \quad [12]$$

A value greater than zero indicates a redundancy-dominated dynamic, while a value less than zero indicates a synergy dominated dynamic.

References

1. NM Timme, et al., High-Degree Neurons Feed Cortical Computations. PLOS Comput. Biol. **12**, e1004858 (2016) Number: 5 Publisher: Public Library of Science.
2. JA Michaels, B Dann, RW Intveld, H Scherberger, Predicting Reaction Time from the Neural State Space of the Premotor and Parietal Grasping Network. J. Neurosci. **35**, 11415–11432 (2015) Publisher: Society for Neuroscience Section: Articles.
3. B Dann, JA Michaels, S Schaffelhofer, H Scherberger, Uniting functional network topology and oscillations in the fronto-parietal single unit network of behaving primates. eLife **5**, e15719 (2016) Publisher: eLife Sciences Publications, Ltd.
4. PG Musial, SN Baker, GL Gerstein, EA King, JG Keating, Signal-to-noise ratio improvement in multiple electrode recording. J. Neurosci. Methods **115**, 29–43 (2002).

5. RQ Quiroga, Z Nadasdy, Y Ben-Shaul, Unsupervised Spike Detection and Sorting with Wavelets and Superparamagnetic Clustering. Neural Comput. **16**, 1661–1687 (2004).
6. A Kraskov, N Dancause, MM Quallo, S Shepherd, RN Lemon, Corticospinal Neurons in Macaque Ventral Premotor Cortex with Mirror Properties: A Potential Mechanism for Action Suppression? Neuron **64**, 922–930 (2009).
7. SN Gozani, JP Miller, Optimal discrimination and classification of neuronal action potential waveforms from multiunit, multichannel recordings using software-based linear filters. IEEE Transactions on Biomed. Eng. **41**, 358–372 (1994) Conference Name: IEEE Transactions on Biomedical Engineering.
8. AD Savva, GD Mitsis, GK Matsopoulos, Assessment of dynamic functional connectivity in resting-state fMRI using the sliding window technique. Brain Behav. **9** (2019).
9. P Wollstadt, et al., IDTxl: The Information Dynamics Toolkit xl: a Python package for the efficient analysis of multivariate information dynamics in networks. J. Open Source Softw. **4**, 1081 (2019) Number: 34.
10. SP Faber, NM Timme, JM Beggs, EL Newman, Computation is concentrated in rich clubs of local cortical networks. Netw. Neurosci. pp. 1–21 (2018).
11. JT Lizier, The Local Information Dynamics of Distributed Computation in Complex Systems, Springer Theses. (Springer Berlin Heidelberg, Berlin, Heidelberg), (2013).
12. M Wibral, J Lizier, S Vögler, V Priesemann, R Galuske, Local active information storage as a tool to understand distributed neural information processing. Front. Neuroinformatics **8** (2014) Publisher: Frontiers.
13. F Takens, Detecting strange attractors in turbulence in Dynamical Systems and Turbulence, Warwick 1980, eds. D Rand, LS Young. (Springer Berlin Heidelberg, Berlin, Heidelberg) Vol. 898, pp. 366–381 (1981) Series Title: Lecture Notes in Mathematics.
14. A Mason, A Nicoll, K Stratford, Synaptic transmission between individual pyramidal neurons of the rat visual cortex in vitro. The J. Neurosci. The Off. J. Soc. for Neurosci. **11**, 72–84 (1991).
15. P Barthó, et al., Characterization of Neocortical Principal Cells and Interneurons by Network Interactions and Extracellular Features. J. Neurophysiol. **92**, 600–608 (2004) Publisher: American Physiological Society.
16. L Novelli, P Wollstadt, P Mediano, M Wibral, JT Lizier, Large-scale directed network inference with multivariate transfer entropy and hierarchical statistical testing. Netw. Neurosci. **3**, 827–847 (2019) Number: 3.
17. SP Sherrill, NM Timme, JM Beggs, EL Newman, Correlated activity favors synergistic processing in local cortical networks in vitro at synaptically relevant timescales. Netw. Neurosci. (Cambridge, Mass.) **4**, 678–697 (2020).
18. T Schreiber, Measuring Information Transfer. Phys. Rev. Lett. **85**, 461–464 (2000) Number: 2.
19. CJ Quinn, TP Coleman, N Kiyavash, NG Hatsopoulos, Estimating the directed information to infer causal relationships in ensemble neural spike train recordings. J. Comput. Neurosci. **30**, 17–44 (2011).
20. S Ito, et al., Large-Scale, High-Resolution Multielectrode-Array Recording Depicts Functional Network Differences of Cortical and Hippocampal Cultures. PLOS ONE **9**, e105324 (2014) Number: 8.
21. M Wibral, R Vicente, M Lindner, Transfer Entropy in Neuroscience in Directed Information Measures in Neuroscience, Understanding Complex Systems, eds. M Wibral, R Vicente, JT Lizier. (Springer, Berlin, Heidelberg), pp. 3–36 (2014).
22. NM Timme, C Lapish, A Tutorial for Information Theory in Neuroscience. eNeuro **5**, ENEURO.0052–18.2018 (2018) Number: 3.
23. S Stramaglia, T Scagliarini, BC Daniels, D Marinazzo, Quantifying Dynamical High-Order Interdependencies From the O-Information: An Application to Neural Spiking Dynamics. Front. Physiol. **11** (2021) Publisher: Frontiers.
24. S Nigam, et al., Rich-Club Organization in Effective Connectivity among Cortical Neurons. J. Neurosci. **36**, 670–684 (2016) Publisher: Society for Neuroscience Section: Articles.
25. N Timme, et al., Multiplex Networks of Cortical and Hippocampal Neurons Revealed at Different Timescales. PLOS ONE **9**, e115764 (2014) Publisher: Public Library of Science.
26. M Shimono, JM Beggs, Functional Clusters, Hubs, and Communities in the Cortical Microconnectome. Cereb. Cortex **25**, 3743–3757 (2015).
27. L Novelli, JT Lizier, Inferring network properties from time series using transfer entropy and mutual information: Validation of multivariate versus bivariate approaches. Netw. Neurosci. **5**, 373–404 (2021).
28. M Ursino, G Ricci, E Magosso, Transfer Entropy as a Measure of Brain Connectivity: A Critical Analysis With the Help of Neural Mass Models. Front. Comput. Neurosci. **14** (2020) Publisher: Frontiers.
29. N Timme, W Alford, B Flecker, JM Beggs, Synergy, redundancy, and multivariate information measures: an experimentalist's perspective. J. Comput. Neurosci. **36**, 119–140 (2014) Number: 2.
30. A Hagberg, D Schult, P Swart, Exploring Network Structure, Dynamics, and Function using NetworkX. (2008).
31. V Colizza, A Flammini, MA Serrano, A Vespignani, Detecting rich-club ordering in complex networks. Nat. Phys. **2**, 110–115 (2006).
32. MPvd Heuvel, RS Kahn, J Goñi, O Sporns, High-cost, high-capacity backbone for global brain communication. Proc. Natl. Acad. Sci. **109**, 11372–11377 (2012) Publisher: National Academy of Sciences Section: Biological Sciences.
33. MA Smith, A Kohn, Spatial and Temporal Scales of Neuronal Correlation in Primary Visual Cortex. J. Neurosci. **28**, 12591–12603 (2008) Publisher: Society for Neuroscience Section: Articles.
34. H Markram, et al., Reconstruction and Simulation of Neocortical Microcircuitry. Cell **163**, 456–492 (2015) Publisher: Elsevier.
35. S Gerhard, et al., The Connectome Viewer Toolkit: An Open Source Framework to Manage, Analyze, and Visualize

Connectomes. Front. Neuroinformatics **0** (2011) Publisher: Frontiers.
36. R Perin, TK Berger, H Markram, A synaptic organizing principle for cortical neuronal groups. Proc. Natl. Acad. Sci. **108**, 5419–5424 (2011) Publisher: National Academy of Sciences Section: Biological Sciences.
37. E Maris, R Oostenveld, Nonparametric statistical testing of EEG- and MEG-data. J. Neurosci. Methods **164**, 177–190 (2007).
38. LGS Jeub, O Sporns, S Fortunato, Multiresolution Consensus Clustering in Networks. Sci. Reports **8**, 3259 (2018).
39. F Rosas, PAM Mediano, M Gastpar, HJ Jensen, Quantifying High-order Interdependencies via Multivariate Extensions of the Mutual Information. Phys. Rev. E **100**, 032305 (2019) Number: 3 arXiv: 1902.11239.
40. T Varley, E Hoel, Emergence as the conversion of information: A unifying theory. arXiv:2104.13368 [cs, math] (2021) arXiv: 2104.13368.