Supporting Information

Broken detailed balance and entropy production in the human brain

Christopher W. Lynn[1,2], Eli J. Cornblath[3,4], Lia Papadopoulos[5], Maxwell A. Bertolero[4], and Danielle S. Bassett[3,6,7,8,9,10,*]

[1]*Initiative for the Theoretical Sciences, Graduate Center, City University of New York, New York, NY 10016, USA*

[2]*Joseph Henry Laboratories of Physics, Princeton University, Princeton, NJ 08544, USA*

[3]*Department of Bioengineering, School of Engineering & Applied Science, University of Pennsylvania, Philadelphia, PA 19104, USA*

[4]*Department of Neuroscience, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA*

[5]*Institute of Neuroscience, University of Oregon, Eugene, Oregon 97403, USA*

[6]*Department of Electrical & Systems Engineering, School of Engineering & Applied Science, University of Pennsylvania, Philadelphia, PA 19104, USA*

[7]*Department of Physics & Astronomy, College of Arts & Sciences, University of Pennsylvania, Philadelphia, PA 19104, USA*

[8]*Department of Neurology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA*

[9]*Department of Psychiatry, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA*

[10]*Santa Fe Institute, Santa Fe, NM 87501, USA*

Contents

1 Introduction

In this Supporting Information, we provide extended analysis and discussion to support the results presented in the main text. In Sec. 2, we describe how the flux vectors (in Figs. 1*D*-*E* and 2*C* of the main text) are calculated and illustrated. In Sec. 3, we show that principal component analysis (PCA) provides a natural low-dimensional embedding of neural dynamics that we can use to visualize fluxes between brain states. In Sec. 4, we show that, although the brain functions out of equilibrium, it does operate at a steady state. Demonstrating that the brain operates at a non-equilibrium steady state opens the door for future investigations using tools and intuitions that have recently been generalized from equilibrium statistical mechanics.[1–3] In Sec. 5, we show that if one shuffles the order of neural time-series data (thereby destroying the arrow of time), then the fluxes between states vanish and the system returns to equilibrium. In Sec. 6, we establish that estimating entropy production using hierarchical clustering yields two desirable properties. First, because a system's entropy production can only decrease with coarse-graining,[4] in order to establish that a system is fundamentally out of equilibrium, one must simply demonstrate that the coarse-grained entropy production is significantly greater than zero. Second, by defining the clusters hierarchically,[5] we prove that the estimated entropy production becomes more accurate as the number of clusters increases. In Sec. 7, we demonstrate how to choose the number of clusters (or coarse-grained states) k when estimating the entropy production. In Sec. 8, we illustrate the $k = 8$ coarse-grained states analyzed in the main text, and demonstrate that the brain operates at a steady state across all of the cognitive tasks. In Sec. 9, we present the fluxes between coarse-grained states in the neural dynamics as directed networks, which we refer to as flux networks. We demonstrate that these flux networks vary in structure across different cognitive tasks. In Sec. 10, we estimate the entropy production of individual humans and demonstrate that, even at the individual level, physical and cognitive exertion yields increases in broken detailed balance. In Secs. 11-13, we show that the entropy production results in the main text do not depend on the assumption that the neural dynamics are Markovian (Sec. 11), are robust to reasonable variation in the number of coarse-grained states k (Sec. 12), and cannot be explained by head movement within the scanner, variance in the neural time-series, nor the lengths of blocks in different tasks

(Sec. 13). In Sec. 14, we compare our main results against phase-randomized surrogate data. Finally, in Sec. 15, we detail how the neural data was processed.

2 Visualizing flux currents

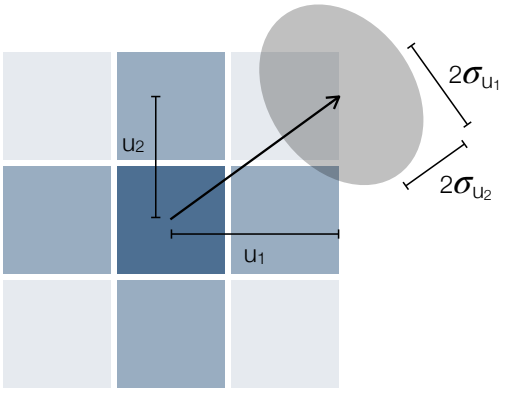
In order to visualize net fluxes in neural dynamics, we project the dynamics onto the first two principal components and employ a technique known as probability flux analysis.[6] The net flux of transitions from a given state (x, y) to its neighboring states can be visualized using the flux vector

$$\boldsymbol{u}(x,y) = \frac{1}{2}\begin{pmatrix} \omega_{(x-1,y),(x,y)} + \omega_{(x,y),(x+1,y)} \\ \omega_{(x,y-1),(x,y)} + \omega_{(x,y),(x,y+1)} \end{pmatrix}. \tag{S1}$$

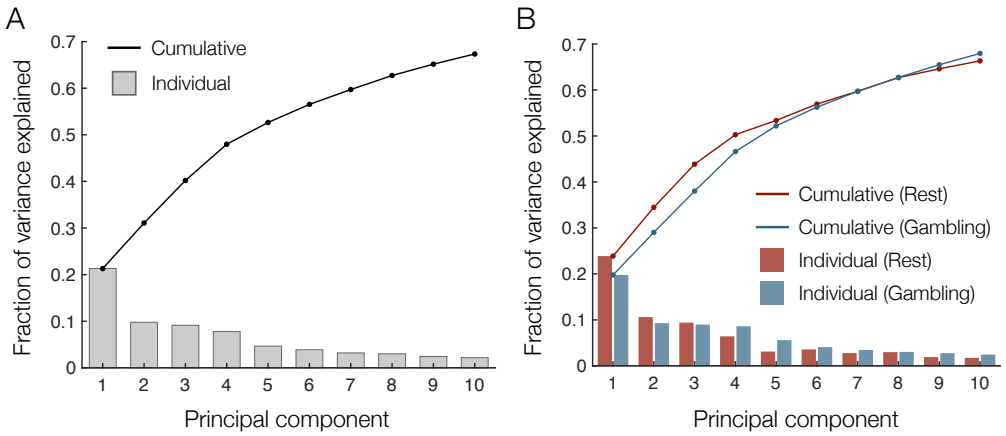
To compute the errors for a given flux vector $\boldsymbol{u}(\boldsymbol{x})$, we calculate the covariance matrix $\mathrm{Cov}(u_1(\boldsymbol{x}), u_2(\boldsymbol{x}))$ by averaging over 100 bootstrapped trajectories (see Materials and Methods in the main text). Then, we illustrate the errors by plotting an ellipse whose axes are aligned with the eigenvectors of the covariance matrix and whose radii are equal to twice the square root of the corresponding eigenvalues (Fig. S1).

3 Low-dimensional embedding using PCA

In order to visualize net fluxes between states in a complex system, we must project the dynamics onto two dimensions. While any pair of dimensions can be used to probe for broken detailed balance, a natural choice is the first two principal components of the time-series data. Indeed, principal



Supporting Fig. 1 | Visualizing flux vectors. Schematic demonstrating how we illustrate the flux of transitions through a state (vector) and the errors in estimating the flux (ellipse).

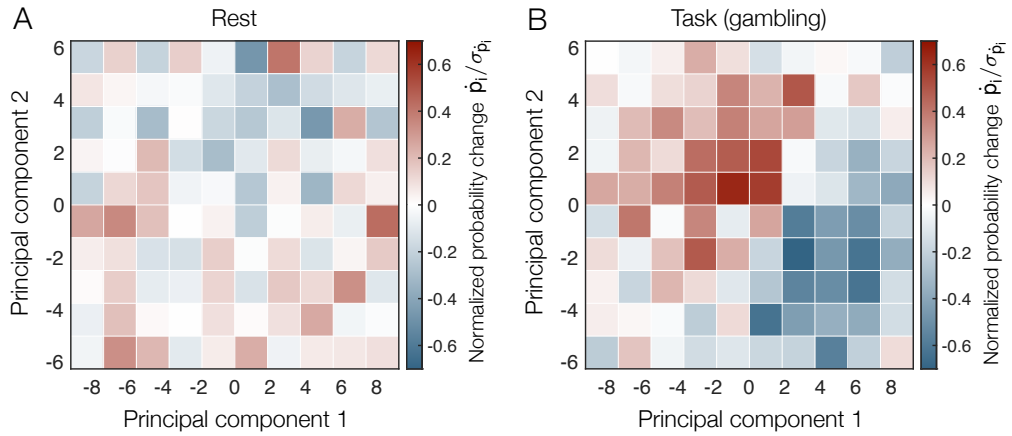


Supporting Fig. 2 | PCA reveals low-dimensional embedding of neural dynamics. (*A*) Cumulative fraction of variance explained by first ten principal components (line) and explained variance for each individual principal component (bars) in the combined rest and gambling data. (*B*) For the same principal components (calculated for the combined rest and gambling data), we plot the cumulative fraction of variance explained (lines) and individual explained variance (bars) for the rest (red) and gambling (blue) data.

component analysis has been widely used to uncover low-dimensional embeddings of large-scale neural dynamics.[7,8] Combining the time-series data from the rest and gambling task scans (that is, the data studied in Fig. 1 in the main text), we find that the first two principal components capture over 30% of the total variance in the observed recordings (Fig. S2*A*), thereby comprising a natural choice for two-dimensional projections. Moreover, we confirm that the projected dynamics capture approximately the same amount of variance in both the rest and gambling tasks, confirming that PCA is not overfitting the neural dynamics in one task or another (Fig. S2*B*).

4 The brain operates at a stochastic steady state

In the main text, we established that the brain breaks detailed balance (Figs. 1 and 4). Yet we did not clarify whether this broken detailed balance arises from changes in the state probabilities themselves, or whether these probabilities are time-invariant and therefore the brain operates at a non-equilibrium steady state. Notably, some of the tools and intuitions developed in traditional statistical mechanics to study equilibrium systems have recently been generalized to systems that operate at non-equilibrium steady states.[1] For example, Evans *et al.* generalized the second law



Supporing Fig. 3 | Small changes in state probabilities imply steady-state dynamics. Change in state probabilities $\dot{p}_i$, normalized by the standard deviation $\sigma_{\dot{p}_i}$, plotted as a function of the first two principal components at rest (*A*) and during the gambling task (*B*).

of thermodynamics to non-equilibrium steady-state systems by discovering the (steady state) fluctuation theorem.[2] More recently, Dieterich *et al.* showed that, by mapping their dynamics to an equilibrium system at an effective temperature, some non-equilibrium steady-state systems are governed by a generalization of the fluctuation-dissipation theorem.[3] Thus, it is both interesting and practical to investigate whether the brain operates at a non-equilibrium steady state. We remark that by "non-equilibrium" we refer to the breaking of detailed balance at large scales, not the obvious non-equilibrium nature of the brain at the cellular and molecular scales.

Previous analyses of fMRI time-series suggest that the brain indeed operates at steady state.[9] To test for steady state in our data, we must examine whether the state probabilities are stationary in time; that is, letting p_i denote the probability of state i, we must determine whether $\dot{p}_i = dp_i/dt = 0$ for all states i. The change in the probability of a state is equal to the net rate at which transitions flow into versus out of a state. For the two-dimensional dynamics studied in Fig. 1 in the main text, this relation takes the form

$$\dot{p}_{(x,y)} = \omega_{(x-1,y),(x,y)} - \omega_{(x,y),(x+1,y)} + \omega_{(x,y-1),(x,y)} - \omega_{(x,y),(x,y+1)}, \tag{S2}$$

where $\omega_{ij} = (n_{ij} - n_{ji})/t_{\text{tot}}$ is the flux rate from state i to state j, n_{ij} is the number of observed transitions $i \to j$, and t_{tot} is the temporal duration of the time-series.[6]

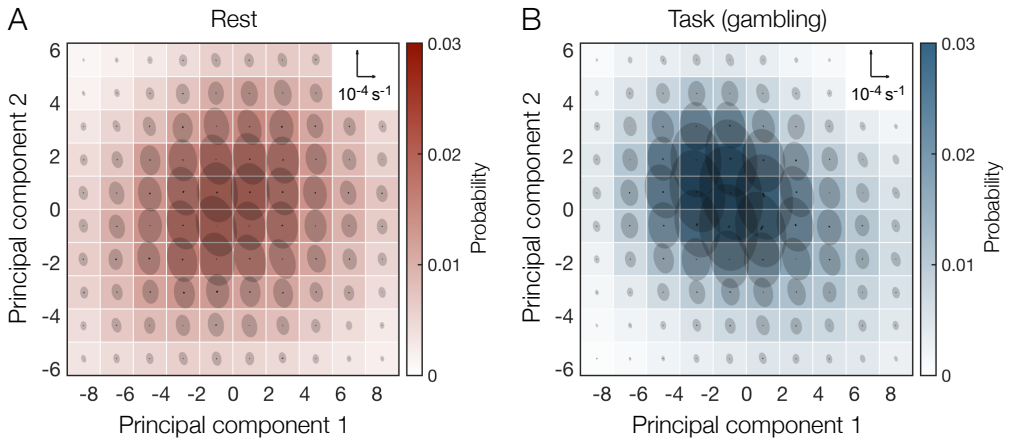
Here, we calculate the changes in state probabilities for both the rest and gambling scans. Across all states in both task conditions, we find that these changes are indistinguishable from zero when compared to statistical noise (Fig. S3). Specifically, the changes in state probabilities are much less than twice their standard deviations, indicating that they cannot be significantly distinguished from zero with a p-value less than 0.05. In Sec. 8, we confirm the same result (that the brain operates at steady-state) for the coarse-grained states analyzed in Fig. 4 in the main text. Combined with the findings in the main text, the stationarity of the neural state probabilities demonstrates that the brain operates at a non-equilibrium steady-state.

5 Shuffling time-series restores detailed balance

In Fig. 1 in the main text, we demonstrate that the brain breaks detailed balance by exhibiting net fluxes between states. Here we demonstrate that if the temporal order of the neural data is destroyed (say, by shuffling the time-series), then the fluxes vanish and detailed balance is restored. Specifically, for both the rest and gambling task scans, we generate 100 surrogate time-series with the order of the data randomly shuffled. Averaging across these shuffled time-series, we find that the fluxes between states are vanishingly small compared to statistical noise (Fig. S4), thus illustrating that the system has returned to detailed balance. We remark that other common surrogate data techniques, such as the random phases and amplitude adjusted Fourier transform surrogates, are not applicable here because they preserve the temporal structure of the time-series data.[10]

6 Estimating entropy production using hierarchical clustering

Complex systems are often high-dimensional, with the number of possible states or configurations growing exponentially with the size of the system. In order to estimate the information entropy production of a complex system (Eq. 1 in the main text), we must reduce the number of states through the use of coarse-graining, or dimensionality reduction, techniques. Interestingly, the entropy production admits a number of strong properties under coarse-graining.[4,11–13] Of particular interest is the fact that the entropy production can only decrease under coarse-graining.[4] Specifically, given two descriptions of a system, a "microscopic" description with states $\{i\}$ and a "macroscopic"

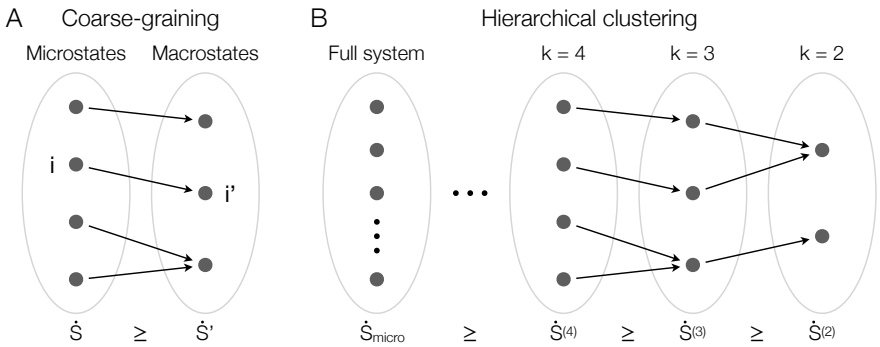


Supporting Fig. 4 | Shuffled data do not exhibit fluxes between neural states. Probability distribution (color) and nearly imperceivable fluxes between states (arrows) for neural dynamics, which are shuffled and projected onto the first two principal components, both at rest (*A*) and during a gambling task (*B*). The flux scale is indicated in the upper right, and the disks represent two-standard-deviation confidence intervals that arise due to finite data (see Materials and Methods in the main text).

description with states $\{i'\}$, we say that the second description is a coarse-graining of the first if there exists a surjective map from the microstates $\{i\}$ to the macrosctates $\{i'\}$ (that is, if each microstate i maps to a unique macrostate i'; Fig. S5*A*). Given such a coarse-graining, Esposito showed[4] that the entropy production of the macroscopic description $\dot{S}'$ can be no larger than that of the microscopic description $\dot{S}$; in other words, the coarse-grained entropy production provides a lower bound for the original value, such that $\dot{S}' \leq \dot{S}$.

The monotonic decrease of the entropy production under coarse-graining implies two desirable mathematical results. First, if one finds that any coarse-grained description of a system breaks detailed balance (that is, if the entropy production at any level of coarse-graining is significantly greater than zero), then one has immediately established that the full microscopic system breaks detailed balance. Thus, even without knowledge of the microscopic non-equilibrium processes at play, one can establish that the brain fundamentally breaks detailed balance at small scales simply by observing violations of detailed balance at large scales (Figs. 1 and 4 in the main text).

By extending this logic, here we show that hierarchical clustering provides systematic im-



Supporting Fig. 5 | Hierarchy of lower bounds on entropy production. (*A*) Coarse-graining is defined by a surjective map from a set of microstates $\{i\}$ to a set of macrostates $\{i'\}$. Under coarse-graining the entropy production cannot increase. (*B*) In hierarchical clustering, states are iteratively combined to form new coarse-grained states (or clusters). Each iteration defines a coarse-graining from k states to $k-1$ states, thereby forming a hierarchy of lower bounds on the entropy production.

provements to the entropy production estimates. In hierarchical clustering, each cluster (or coarse-grained state) at one level of description (with k clusters) maps to a unique cluster at the level below (with $k-1$ clusters; Fig. S5*B*). This process can either be carried out by starting with a large number of clusters and then iteratively picking pairs of clusters to combine (known as agglomerative clustering), or by starting with a small number of clusters and then iteratively picking one cluster to split into two (known as divisive clustering, which we employ in our analysis).[14] In both cases, the mapping from k clusters to $k-1$ clusters is surjective, thereby representing a coarse-graining of the system, as defined previously. Thus, letting $\dot{S}^{(k)}$ denote the entropy production estimated with k clusters, hierarchical clustering defines a hierarchy of lower bounds on the microscopic entropy production $\dot{S}_{\text{micro}}$:

$$0 = \dot{S}^{(1)} \leq \dot{S}^{(2)} \leq \dot{S}^{(3)} \leq \ldots \leq \dot{S}_{\text{micro}}. \tag{S3}$$

This hierarchy, in turn, demonstrates that the estimated entropy production $\dot{S}^{(k)}$ becomes larger (and thus more accurate) with increasing k.

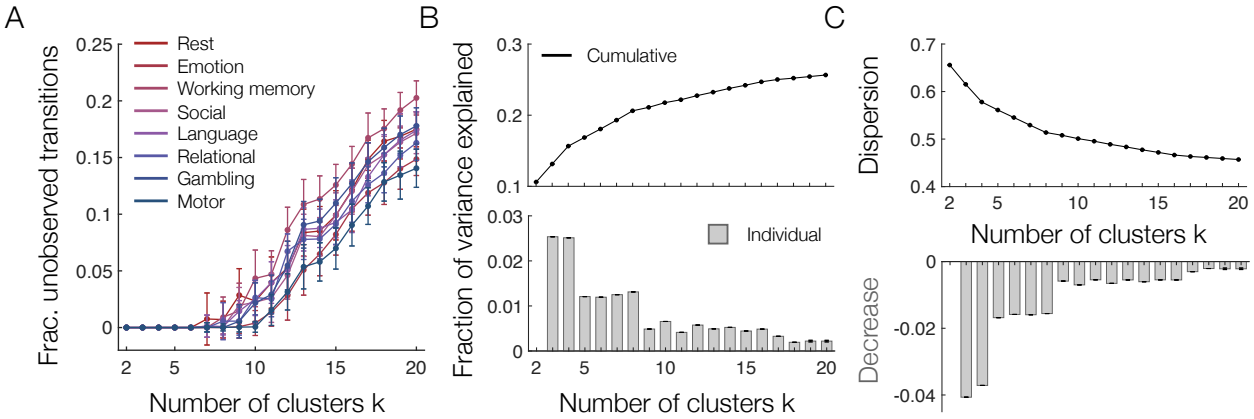
We remark that the discussion above neglects finite data effects. We recall that estimating the entropy production requires first estimating the transition probabilities P_{ij} from state i to state j. This means that for Markov systems with k clusters, one must estimate k^2 different probabilities.

Thus, while increasing k improves the accuracy of the estimated entropy production in theory, in practice increasing k eventually leads to sampling issues that decrease the accuracy of the estimate. Given these competing influences, when analyzing real data the goal should be to choose k such that it is as large as possible while still providing accurate estimates of the transition probabilities. We discuss a systematic strategy for choosing k in the following section.

7 Choosing the number of coarse-grained states

As discussed above, when calculating the entropy production, we wish to choose a number of coarse-grained states k that is as large as possible while still arriving at an accurate estimate of the transition probabilities. One simple condition for estimating each transition probability P_{ij} is that we observe the transition $i \to j$ at least once in the time-series. For all of the different tasks, Fig. S6*A* shows the fraction of the k^2 state transitions that are left unobserved after coarse-graining with k clusters. We find that $k = 8$ is the largest number of clusters for which the fraction of unobserved transitions equals zero (within statistical errors) for all tasks; that is, the largest number of clusters for which all state transitions across all tasks were observed at least once. For this reason, we use $k = 8$ coarse-grained states to analyze the brain's entropy production (Fig. 4 in the main text).

Interestingly, we find that $k = 8$ coarse-grained states is a good choice for two additional reasons. The first comes from studying the amount of variance explained by k clusters (Fig. S6*B*). We find that the increase in explained variance from $k - 1$ to k clusters is roughly constant for $k = 3$ and 4, then $k = 5$ to 8, and then $k = 9$ to 16. This pattern means that $k = 4$, 8, and 16 are natural choices for the number of coarse-grained states, since any further increase (say from $k = 8$ to 9) will yield a smaller improvement in explained variance. Similarly, the second reason for choosing $k = 8$ comes from studying the average distance between states within a cluster, which is known as the dispersion (Fig. S6*C*). Intuitively, a coarse-grained description with low dispersion provides a good fit of the observed data. Similar to the explained variance, we find that the decrease in dispersion from $k - 1$ to k clusters is nearly constant for $k = 3$ to 4, then $k = 5$ to 8, and then $k = 9$ to 16, once again suggesting that $k = 4$, 8, and 16 are natural choices for the number of clusters. Together, these results demonstrate that the coarse-grained description with



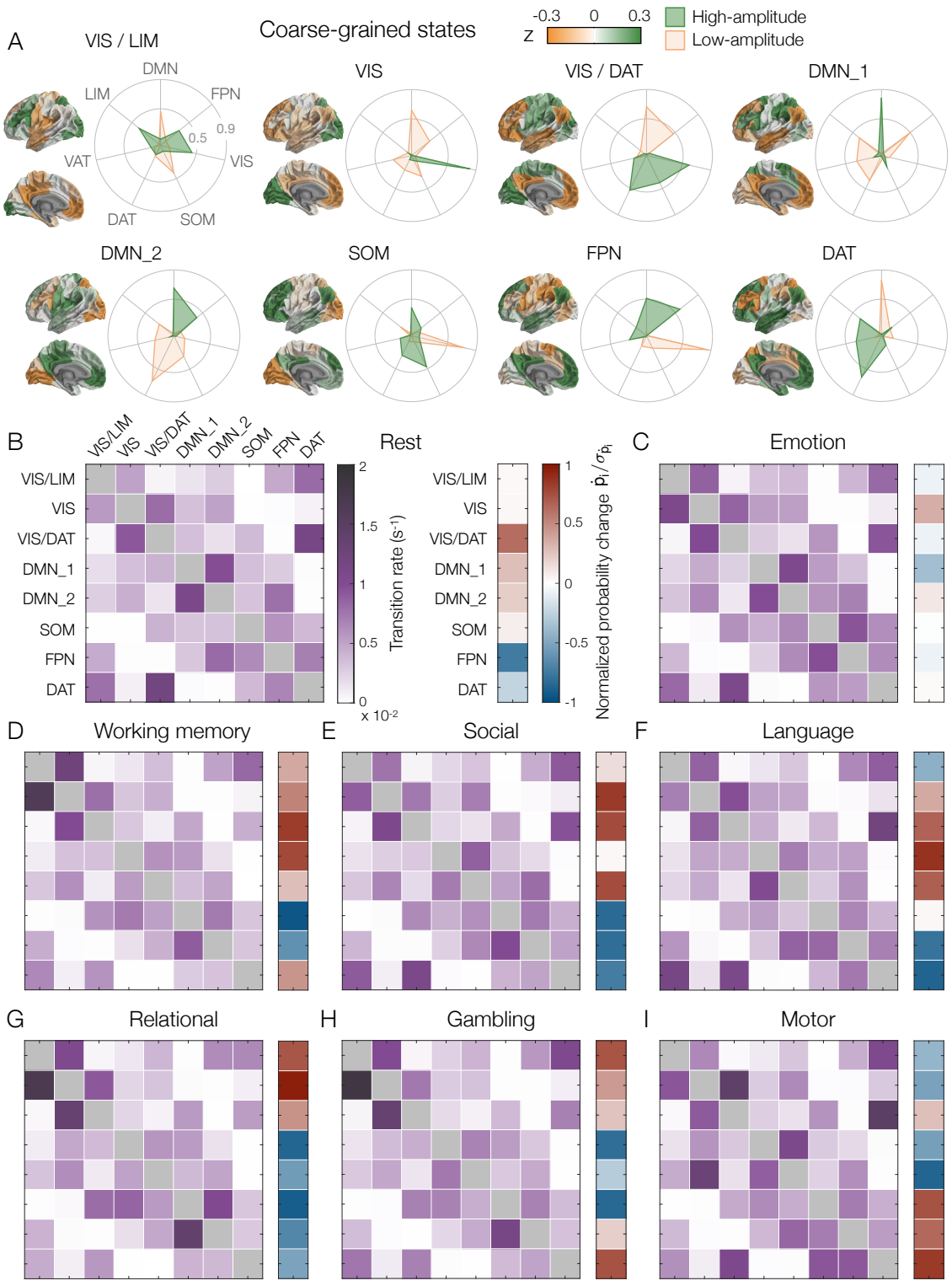
Supporting Fig. 6 | Choosing the number of coarse-grained states k. (*A*) Fraction of the k^2 state transitions that remain unobserved after hierarchical clustering with k clusters for the different tasks. Error bars represent two-standard-deviation confidence intervals that arise due to finite data. (*B*) Percent variance explained (top) and the increase in explained variance from $k-1$ to k clusters (bottom) as functions of k. (*C*) Dispersion, or the average distance between data points within a cluster (top), and the decrease in dispersion from $k-1$ to k clusters (bottom) as functions of k.

$k = 8$ states provides a good fit to the neural time-series data while still allowing for an accurate estimate of the entropy production in each task.

8 Coarse-grained states and transition matrices

We are now prepared to examine the $k = 8$ coarse-grained neural states and the transitions between them. In Fig. S7*A*, we project each of the eight coarse-grained states onto seven pre-defined cognitive systems.[15] We find that each of the states corresponds to high-amplitude activity in one or two systems, and we label each state based on this high-amplitude activity.

We can now measure the rate of transitions between the coarse-grained states during different cognitive tasks (Fig. S7*B-I, left*). We note that the transition rates are given simply by dividing the transition probabilities P_{ij} by the fMRI repetition time $\Delta t = 0.72$ s. At rest, we find that the transition rates appear nearly symmetric, indicating that the brain is close to detailed balance (Fig. S7*B, left*). By contrast, careful examination of the transition rates during the seven cognitive tasks reveals asymmetries, illustrating broken detailed balance (Fig. S7*C-I, left*).



Supporting Fig. 7 | Transitions between coarse-grained neural states. (*A*) Coarse-grained neural states calculated using hierarchical clustering ($k = 8$), with surface plots indicating the

z-scored activation of different brain regions. For each state, we calculate the cosine similarity between its high-amplitude (green) and low-amplitude (orange) components and seven pre-defined neural systems:[15] default mode (DMN), frontoparietal (FPN), visual (VIS), somatomotor (SOM), dorsal attention (DAT), ventral attention (VAT), and limbic (LIM). We label each state according to its largest high-amplitude cosine similarities. (*B-I*) Transition rates (*left*) between each of the eight coarse-grained states and the changes in state probabilities $\dot{p}_i$ normalized by their standard deviations $\sigma_{\dot{p}_i}$ (*right*) at rest (*B*) and during seven cognitive tasks: emotional processing (*C*), working memory (*D*), social inference (*E*), language processing (*F*), relational matching (*G*), gambling (*H*), and motor execution (*I*). Transition rates are computed by dividing the transition probabilities P_{ij} by the fMRI repetition time $\Delta t = 0.72$ s. Across all panels, averages and standard deviations are computed over 100 bootstrap samples.

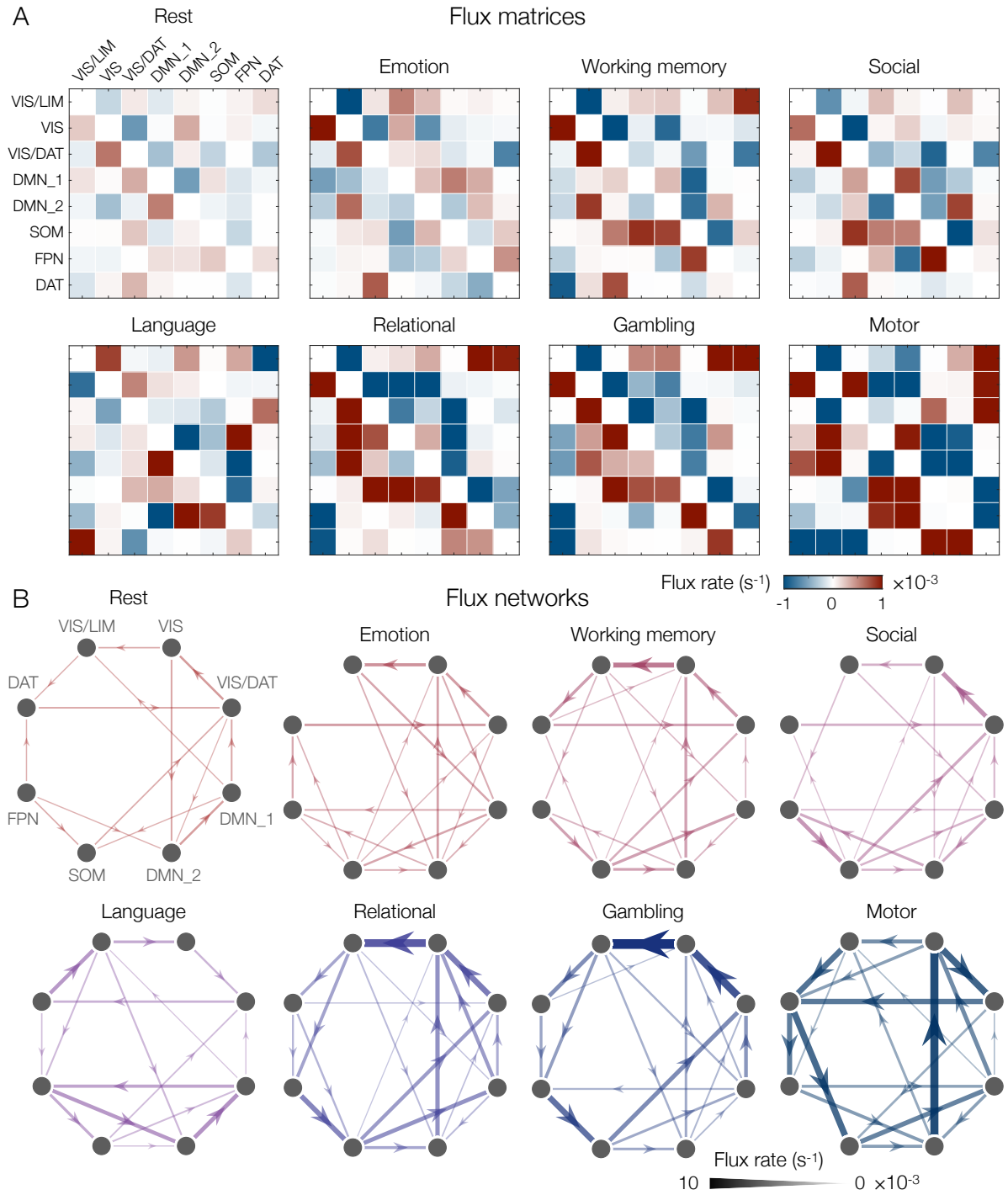
Given the transition rates, we can now determine whether the brain operates at steady state. We note that the rate of change of the state probability p_i is given by

$$\dot{p}_i = \frac{1}{\Delta t} \sum_j \left(P_{ji} - P_{ij} \right). \tag{S4}$$

Computing the change $\dot{p}_i$ in the probability of each state in Fig. S7*A*, and normalizing by the standard deviation $\sigma_{\dot{p}_i}$ (computed over 100 bootstrap samples), we are able to determine whether any of the state probabilities vary significantly in time. Across all states in all tasks, we find that the changes in state probabilities are much less than twice their standard deviations (Fig. S7*B-I, right*). We therefore conclude, at least at this coarse-grained level, that the brain operates at steady state.

9 Flux networks: Visualizing fluxes between coarse-grained states

In Fig. 4 in the main text, we demonstrate that the brain has the capacity to operate at a wide range of distances from detailed balance. We did so by estimating the entropy production of neural dynamics during different cognitive tasks. In addition to investigating the entropy production, one can also examine the specific neural processes underlying the violations of detailed balance, which are encoded in the fluxes between neural states. The rate of flux from state i to state j is given by $P_{ij} - P_{ji}$, normalized by the fMRI repetition time $\Delta t = 0.72$ s. In other words, the fluxes are simply the antisymmetric parts of the transition rates in Fig. S7. We find that these fluxes are small



Supporting Fig. 8 | **Each cognitive task induces a unique pattern of fluxes.** (*A*) Flux rates between the eight coarse-grained states at rest and during the seven cognitive tasks. The flux rate from state i to state j is given by $P_{ij} - P_{ji}$, normalized by the fMRI repetition time $\Delta t = 0.72$ s. (*B*) Flux networks illustrating the fluxes in panel *A*. Edge weights indicate flux

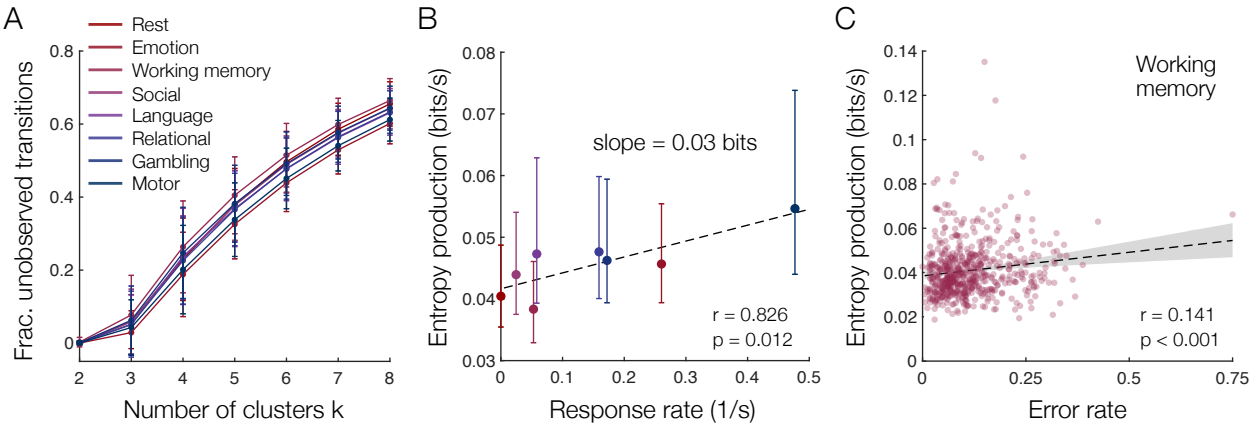
rates, and fluxes are only included if they are significant relative to the noise floor induced by finite data (one-sided *t*-test, $p < 0.001$).

during rest, and much stronger during different cognitive tasks (Fig. S8*A*). Moreover, we find that each task induces a unique pattern of fluxes between neural states.

To visualize the pattern of fluxes, we introduce the concept of a flux network, with nodes representing the coarse-grained states and directed edges reflecting fluxes between states (Fig. S8*B*). These networks illustrate, for example, that the fluxes almost vanish during rest, thereby indicating that the brain nearly obeys detailed balance. Interestingly, in the emotion, working memory, social, relational, and gambling tasks—all of which involve visual stimuli—the strongest fluxes connect visual (VIS) states. By contrast, these fluxes are weak in the language task, which only involves auditory stimuli. Finally, in the motor task, wherein subjects are prompted to execute physical movements, the dorsal attention (DAT) state mediates fluxes between disparate parts of the network, perhaps reflecting the role of the DAT system in directing goal-oriented attention.[16,17] In this way, broken detailed balance in the brain is not driven by a single subsystem, but rather involves different combinations of subsystems depending on the specific task being performed. Examining the structural properties and cognitive neuroscientific interpretations of these flux networks is an important direction for future studies.

10 Entropy production in individual humans

Throughout the main text, we combine the neural dynamics of all subjects in order to increase the statistical power of our analyses. However, it is also interesting to investigate violations of detailed balance in individual humans. The primary difficulty in doing so lies in estimating the transition probabilities P_{ij} that are required to compute the entropy production (Eq. 1 in the main text). As discussed in Sec. 7, concatenating the neural time-series across subjects allows us to estimate the transition probabilities using $k = 8$ coarse-grained states (Fig. S6). By contrast, when analyzing individual subjects, the largest number of states for which we observe every transition at least once in each task is $k = 3$ (Fig. S9*A*).



Supporting Fig. 9 | Entropy production in individual subjects. (*A*) Fraction of the k^2 state transitions that remain unobserved after hierarchical clustering with k clusters for each subject and each task. Data points and error bars represent two standard deviations over the 590 different subjects. (*B*) Entropy production of individual subjects increases with the rate of physical responses across the different task settings (Pearson correlation $r = 0.826$, $p = 0.012$). Data points and error bars represent medians and interquartile ranges over the 590 subjects, and the dashed line indicates linear best fit. (*C*) Entropy production increases with error rate in the working memory task (Pearson correlation $r = 0.141$, $p < 0.001$). We confirm that relationship also holds after removing outliers (Pearson correlation $r = 0.133$, $p = 0.002$). Dashed line and shaded region indicate linear best fit and 95% confidence interval. To account for finite-data errors, all data points in all panels are averaged over 100 bootstrap samples for each subject and each task.

Performing hierarchical clustering with $k = 3$ clusters, we estimate the entropy production for each of the 590 subjects during each task (and rest). As in the main text (Fig. 4), we then investigate the dependence of entropy production on physical and cognitive exertion. Across all tasks, we find that the entropy production of neural dynamics increases significantly with the rate of motor responses (Fig. S9*B*). This result confirms that the population-level relationship between broken detailed balance and physical effort (Fig. 4*B* in the main text) extends to the scale of individual humans.

To examine the relationship between entropy production and cognitive demand, we once again focus on the working memory task. At the population level, we found that the high cognitive

load condition induces a two-fold increase in entropy production over the low load condition (Fig. 4*C* in the main text). However, performing the same analysis on individual subjects is infeasible, since it requires estimating the entropy production for each subject on only a fraction of the working memory data. Instead, as a proxy for cognitive effort, we can examine the rate at which subjects make errors. Indeed, across subjects, we find that entropy production increases significantly with error rate (Fig. S9*C*), confirming that the association between broken detailed balance and cognitive effort persists at the individual level. Together, these results indicate that, even for individual humans, violations of detailed balance grow with physical exertion and cognitive demand.

11 Testing the Markov assumption

Thus far, we have employed a definition of entropy production (Eq. 1 in the main text) that relies on the assumption that the time-series is Markovian; that is, that the state x_t of the system at time t depends only on the previous state x_{t-1} at time $t-1$. For real time-series data, however, the dynamics may not be Markovian, and Eq. 1 in the main text is not exact. In general, the entropy production (per trial) is given by[13,18]

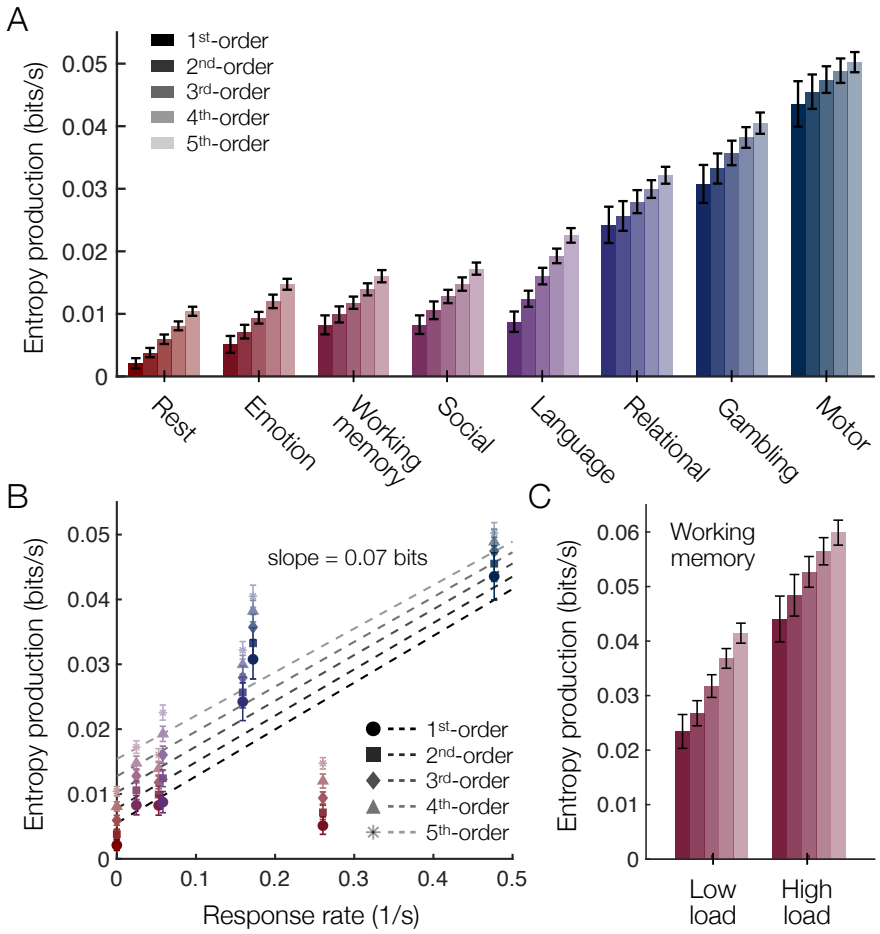
$$\dot{S} = \lim_{t\to\infty} \frac{1}{t} \sum_{i_1,\ldots,i_{t+1}} P_{i_1,\ldots,i_{t+1}} \log \frac{P_{i_1,\ldots,i_{t+1}}}{P_{i_{t+1},\ldots,i_1}}, \tag{S5}$$

where $P_{i_1,\ldots,i_{t+1}}$ is the probability of observing the sequence of states $i_1,\ldots,i_{t+1}$. If the dynamics are Markovian of order ℓ, then Eq. S5 is equivalent to

$$\dot{S} = \frac{1}{\ell} \sum_{i_1,\ldots,i_{\ell+1}} P_{i_1,\ldots,i_{\ell+1}} \log \frac{P_{i_1,\ldots,i_{\ell+1}}}{P_{i_{\ell+1},\ldots,i_1}}. \tag{S6}$$

For example, if $\ell = 1$ we recover the Markov approximation in Eq. 1 in the main text. In general, computing the ℓ^{th}-order entropy production for a system with k states requires estimating the probabilities of all $k^{\ell+1}$ sequences of length $\ell+1$. Thus, the number of independent statistics that need to be estimated grows exponentially with the order ℓ, often making it infeasible to estimate the entropy production beyond order $\ell = 1$.

Despite the abovementioned limitations, here we estimate the entropy production of neural dynamics up to order $\ell = 5$. In doing so, we demonstrate that our main results (Fig. 4 in the main



Supporting Fig. 10 | Higher-order approximations of entropy production in the brain. (*A*) Entropy productions of orders $\ell = 1, 2, 3, 4, 5$ computed at rest and during seven cognitive tasks. All estimates are based on the same coarse-grained states, computed using hierarchical clustering with $k = 8$ clusters. (*B*) Entropy production estimates as a function of response rate for the tasks listed in panel (*A*). Across all orders $\ell = 1, 2, 3, 4, 5$, each response induces an identical 0.07 bits of produced entropy (within errors, $p < 0.05$). (*C*) Entropy production estimates for low cognitive load and high cognitive load conditions in the working memory task, where low and high loads represent 0-back and 2-back conditions, respectively. In all panels, data points of increasing brightness indicate entropy production estimates of increasing order, and error bars reflect-two standard-deviation confidence intervals that arise due to finite data (see Materials and Methods in the main text).

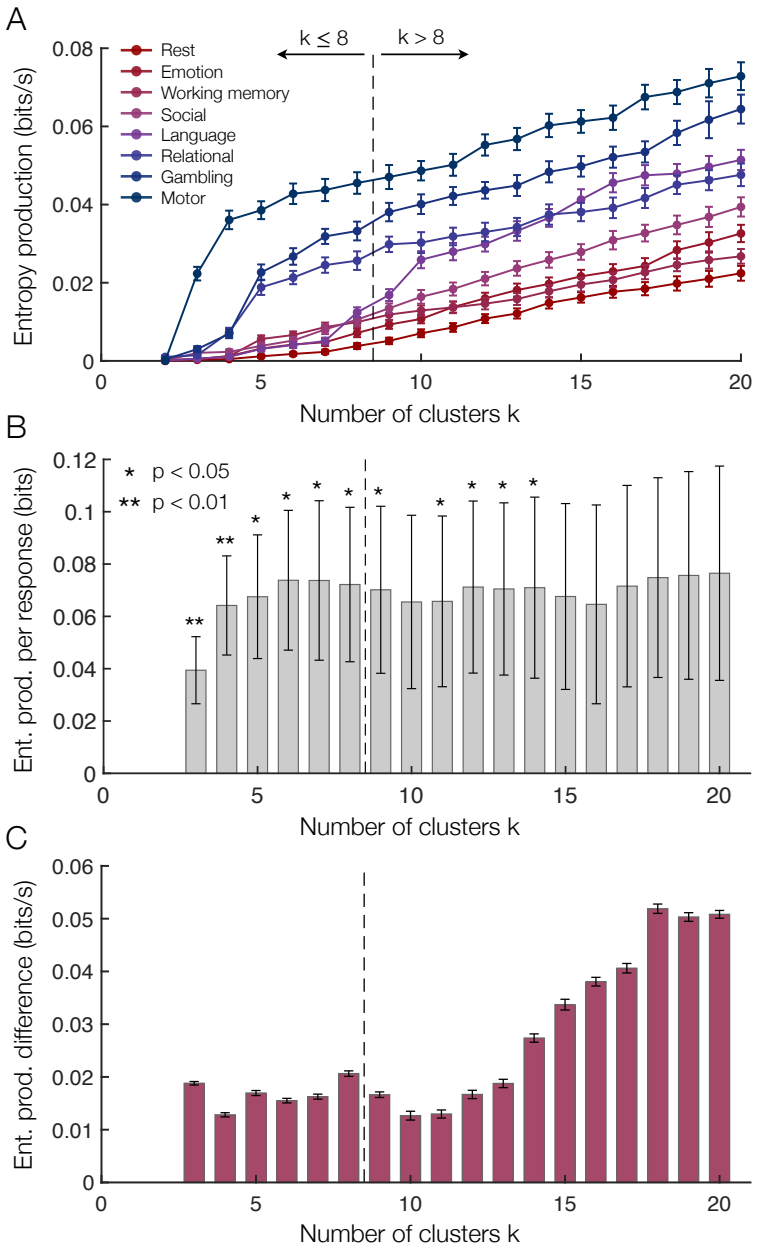
text) do not depend qualitatively on the Markov approximation in Eq. 1 in the main text. Just as we did under the Markov approximation (Fig. 4 in the main text), we cluster the neural data using

$k = 8$ coarse-grained states. Given that we are now required to estimate $k^{\ell+1}$ different probabilities (a number that grows up to 2.6×10^5 for $\ell = 5$) rather than just $k^2 = 64$ probabilities, there are inevitably entries in the sum in Eq. S6 that are infinite (i.e., those corresponding to reverse-time sequences $i_{\ell+1}, \ldots, i_1$ that are not observed in the time-series). As is common,[13,18] we set these terms to zero.

We find that all of the higher-order approximations studied ($\ell = 2, 3, 4, 5$) yield exactly the same hierarchy of entropy productions across task settings (Fig. S10*A*) as the first-order approximation (Fig. 4*A* in the main text). In particular, across all orders ℓ, the neural dynamics produce less entropy during rest than during each of the cognitive tasks (Fig. S10*A*). Moreover, the higher-order entropy productions remain significantly correlated with the frequency of physical responses in different tasks (Fig. S10*B*). In fact, for all orders ℓ examined, each response induces an identical 0.07 bits of produced entropy (within errors; Fig. S10*B*). Finally, in the working memory task, the higher-order entropy productions remain larger in the high cognitive load condition than in the low-load condition (Fig. S10*C*). Specifically, the neural dynamics produce an additional 0.02 bits per second of entropy in the high-load condition compared to the low-load condition, a difference that is identical (within errors) across all of the Markov orders studied. Together, these results demonstrate that the central conclusions of the main text generalize to higher-order Markov approximations.

12 Varying the number of coarse-grained states

In Sec. 7, we presented methods for choosing the number of coarse-grained states k, concluding that $k = 8$ is an appropriate choice for our neural data. However, it is important to check that the entropy production results from Fig. 4 in the main text do not vary significantly with our choice of k. In Fig. S11*A*, we plot the estimated entropy production for each task setting (including rest) as a function of the number of coarse-grained states k. We find that the tasks maintain approximately the same ordering across all choices of k considered, with the brain producing the least entropy during rest, the most entropy during the motor task, and the second most entropy during the gambling task. Furthermore, we find that the correlation between entropy production and physical



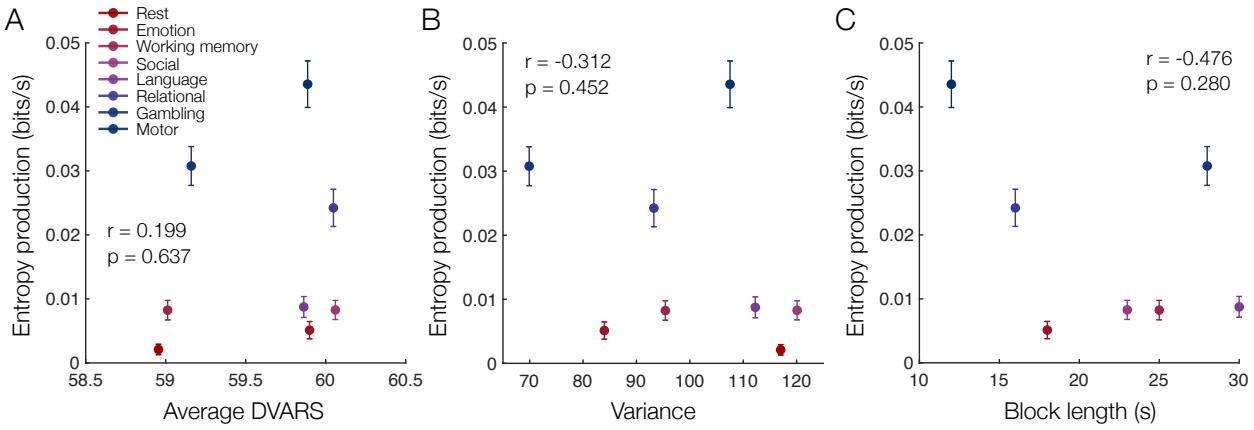
Supporting Fig. 11 | Entropy production in the brain at different levels of coarse-graining. (*A*) Entropy production at rest and during seven cognitive tasks as a function of the number of clusters k used in the hierarchical clustering. Error bars reflect two-standard-deviation confidence intervals that arise due to finite data (see Materials and Methods in the main text). (*B*) Slope of the linear relationship between entropy production and physical response rate across tasks for different numbers of clusters k. Error bars represent one-standard-deviation confidence intervals of the slope and asterisks indicate the significance of the Pearson correlation between entropy production and response rate. (*C*) Difference bet-

ween the entropy production during high-load and low-load conditions of the working memory task as a function of the number of clusters k. Error bars represent two-standard-deviation confidence intervals that arise due to finite data (see Materials and Methods in the main text), and the entropy production difference is significant across all values of k (one-sided *t*-test, $p < 0.001$).

response rate (Fig. 4*B* in the main text) remains significant for all $k \leq 8$ (that is, for all choices of k for which we observe all transitions at least once in each task; Fig. S6*A*) as well as $k = 9$, 11, 12, 13, and 14 (Fig. S11*B*). We remark that we do not study the case $k = 2$ because the entropy production is zero by definition for steady-state systems with two states (Fig. S11*A*). Finally, we confirm that the entropy production is significantly larger during high-cognitive-load conditions than low-cognitive-load conditions in the working memory task (Fig. 4*C* in the main text) for all choices of k considered (Fig. S11*C*). Together, these results demonstrate that the relationships between entropy production and physical and cognitive effort are robust to reasonable variation in the number of coarse-grained states k.

13 Robustness to head motion, signal variance, and block length

Here, we show that the effects of physical response rate and cognitive effort on entropy production cannot be explained by head movement within the scanner (a common confound in fMRI studies[19]), variance in the neural time-series, nor the block lengths of different tasks. To quantify head movement, for each time point in every time-series, we compute the spatial standard deviation of the difference between the current image and the previous image. This quantity, known as DVARS, is a common measure of head movement in fMRI data.[20] Importantly, we find that entropy production is not significantly correlated with the average DVARS within each task (Fig. S12*A*), thereby demonstrating that the relationship between entropy production and physical response rate is not simply due to the confound of subject head movement within the scanner. Additionally, we find that entropy production is not significantly correlated with the variance of the neural data within each task (Fig. S12*B*). We therefore establish that our entropy production estimates are not simply driven by variations in the amount of noise in the neural data across different tasks.



Supporting Fig. 12 | Entropy production in the brain cannot be explained by head movement, signal variance, nor block length. Entropy production versus the average DVARS (*A*), the variance of the neural time-series (*B*), and the lengths of task blocks (*C*) at rest and during seven cognitive tasks. Across all panels, entropy productions are estimated using hierarchical clustering with $k = 8$ clusters and are divided by the fMRI repetition time $\Delta t = 0.72$ s to compute entropy production rates. Error bars reflect two-standard-deviation confidence intervals that arise due to finite data (see Materials and Methods in the main text).

Finally, one might suspect that the fluxes between neural states (and therefore the entropy production) is driven by variations in the block structure of different tasks. However, we do not find a significant relationship between the entropy production and the lengths of blocks in different tasks (Fig. S12*C*), thereby indicating that block structure alone cannot explain the observed variations in entropy production across tasks.

14 Comparing against phase-randomized surrogate data

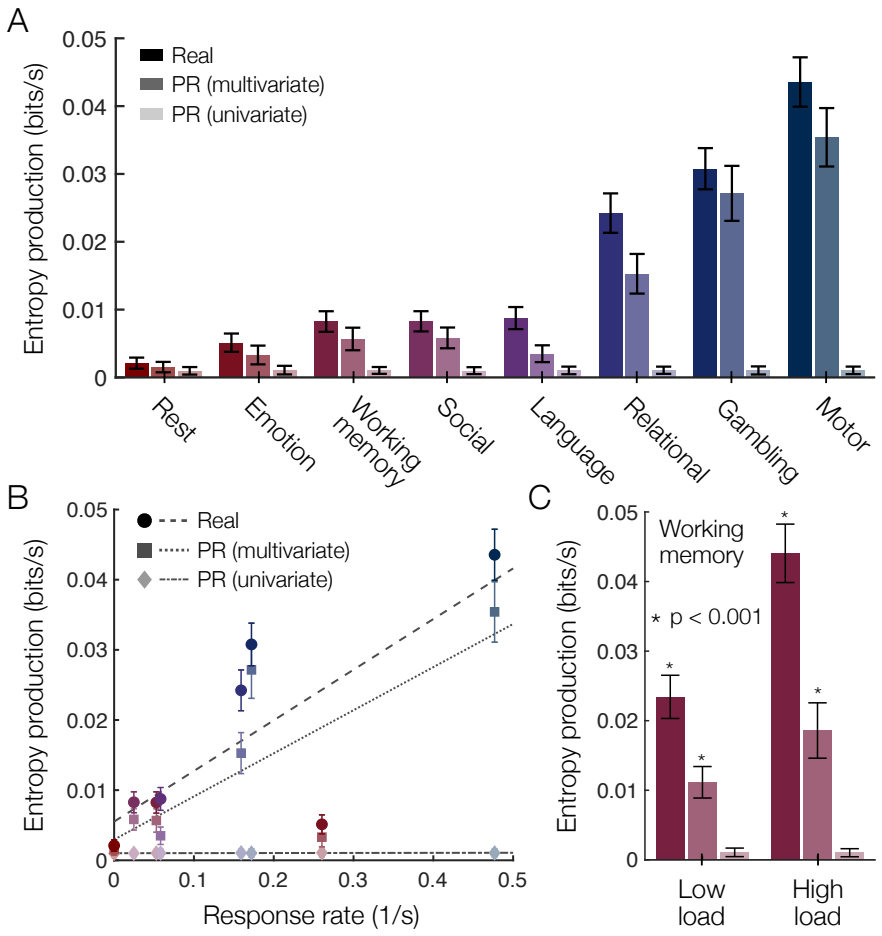
In Sec. 5, we showed that if one shuffles the order of neural time-series data (thereby destroying the arrow of time), then the fluxes between states vanish and the system returns to detailed balance (Fig. S4). Yet it is also interesting to consider the effect of other surrogate data techniques on broken detailed balance. Here, we consider two such techniques: one that randomizes the phases of neural activity while conserving both the auto-correlations and cross-correlations between regions (referred to as the multivariate phase-randomized (MPR) surrogate), and another that randomizes phases while conserving auto-correlations but not cross-correlations (referred to as the univariate

phase-randomized (UPR) surrogate).[10,21] Importantly, the auto-correlations are symmetric in time, while the cross-correlations are not. Therefore, randomizing the data while holding only the auto-correlations fixed (as in the UPR surrogate) should destroy the arrow of time and return the system to detailed balance. By contrast, holding the cross-correlations fixed (as in the MPR surrogate) should ensure that some of the broken detailed balance remains from the original data.

Here, we compute the entropy productions of MPR and UPR surrogate data during the seven cognitive tasks and rest. For the MPR data, we find that the entropy production (and therefore the distance from detailed balance) remains significant, but is reduced relative to the original data (Fig. S13*A*). Meanwhile, for the UPR data, the entropy production nearly vanishes and the data obeys detailed balance (Fig. S13*B*). Indeed, for the MPR data, we still find significant (yet reduced) increases in entropy production with both physical response rate (Fig. S13*B*) and cognitive load (Fig. S13*C*), while we find no such dependencies in the UPR data. Together, these results confirm that randomizing the data while holding the auto-correlations fixed returns the system to detailed balance, whereas holding the cross-correlations fixed preserves some of the broken detailed balance in the original data.

15 Data processing

The resting, emotional processing, working memory, social inference, language processing, relational matching, gambling, and motor execution fMRI scans are from the S1200 Human Connectome Project release.[22,23] Brains were normalized to fslr32k via the MSM-AII registration with 100 regions.[24] CompCor, with five principal components from the ventricles and white matter masks, was used to regress out nuisance signals from the time series. Additionally, the 12 detrended motion estimates provided by the Human Connectome Project were regressed out from the regional time series. The mean global signal was removed and then time series were band-pass filtered from 0.009 to 0.08 Hz. Then, frames with greater than 0.2 mm frame-wise displacement or a derivative root mean square (DVARS) above 75 were removed as outliers. We filtered out sessions composed of greater than 50 percent outlier frames, and we only analyzed data from subjects that had all scans remaining after filtering, leaving 590 individuals. The processing pipeline used here



Supporting Fig. 13 | **Entropy production of phase-randomized surrogate data.** (*A*) Entropy productions of the true data (dark shades), multivariate phase-randomized (MPR) data (medium shades), and univariate phase-randomized (UPR) data (light shades) during rest and seven cognitive tasks. (*B*) Entropy production estimates as functions of response rate for the tasks listed in panel (*A*), where lines indicate linear best fits. For the real and MPR data, responses induce significant entropy production: 0.07 bits per response for real data ($p = 0.02$) and 0.06 bits per response for MPR data ($p = 0.03$). For UPR data, we do not find a significant increase in entropy production with responses ($p = 0.12$). (*C*) Entropy production estimates for low cognitive load and high cognitive load conditions in the working memory task, where low and high loads represent 0-back and 2-back conditions, respectively. We find significant differences in entropy production between the low- and high-load conditions for both the true data (dark shade) and MPR data (medium shade), but not the UPR data (light shade). In all panels, estimates are computed using $k = 8$ coarse-grained states, and error bars reflect two standard-deviation confidence intervals computed using 100 bootstrap samples.

has previously been suggested to be ideal for removing false relations between neural dynamics and behavior.[25] Finally, for each subject and each scan, we only analyze the first 176 time points, corresponding to the length of the shortest task (emotional processing); this truncation controls for the possibility of data size affecting comparisons across tasks.

References

1. Seifert, U. Stochastic thermodynamics, fluctuation theorems and molecular machines. *Rep. Prog. Phys.* **75**, 126001 (2012).

2. Evans, D. J., Cohen, E. G. D. & Morriss, G. P. Probability of second law violations in shearing steady states. *Phys. Rev. Lett.* **71**, 2401 (1993).

3. Dieterich, E., Camunas-Soler, J., Ribezzi-Crivellari, M., Seifert, U. & Ritort, F. Single-molecule measurement of the effective temperature in non-equilibrium steady states. *Nat. Phys.* **11**, 971–977 (2015).

4. Esposito, M. Stochastic thermodynamics under coarse graining. *Phys. Rev. E* **85**, 041125 (2012).

5. Jain, A. K., Murty, M. N. & Flynn, P. J. Data clustering: A review. *ACM Comput. Surv.* **31**, 264–323 (1999).

6. Battle, C. *et al.* Broken detailed balance at mesoscopic scales in active biological systems. *Science* **352**, 604–607 (2016).

7. Cunningham, J. P. & Byron, M. Y. Dimensionality reduction for large-scale neural recordings. *Nat. Neurosci.* **17**, 1500 (2014).

8. Song, X., Ji, T. & Wyrwicz, A. M. Baseline drift and physiological noise removal in high field fMRI data using kernel PCA. In *2008 IEEE International Conference on Acoustics, Speech and Signal Processing*, 441–444 (IEEE, 2008).

9. Liegeois, R., Laumann, T. O., Snyder, A. Z., Zhou, J. & Yeo, B. T. Interpreting temporal fluctuations in resting-state functional connectivity MRI. *Neuroimage* **163**, 437–455 (2017).

10. Lancaster, G., Iatsenko, D., Pidde, A., Ticcinelli, V. & Stefanovska, A. Surrogate data for hypothesis testing of physical systems. *Phys. Rep.* **748**, 1–60 (2018).

11. Martínez, I. A., Bisker, G., Horowitz, J. M. & Parrondo, J. M. Inferring broken detailed balance in the absence of observable currents. *Nat. Commun.* **10**, 1–10 (2019).

12. Seifert, U. Entropy production along a stochastic trajectory and an integral fluctuation theorem. *Phys. Rev. Lett.* **95**, 040602 (2005).

13. Roldán, É. & Parrondo, J. M. Estimating dissipation from single stationary trajectories. *Phys. Rev. Lett.* **105**, 150607 (2010).

14. Cohen-Addad, V., Kanade, V., Mallmann-Trenn, F. & Mathieu, C. Hierarchical clustering: Objective functions and algorithms. In *Proceedings of the Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms*, 378–397 (SIAM, 2018).

15. Thomas Yeo, B. *et al.* The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *J. Neurophysiol.* **106**, 1125–1165 (2011).

16. Fox, M. D., Corbetta, M., Snyder, A. Z., Vincent, J. L. & Raichle, M. E. Spontaneous neuronal activity distinguishes human dorsal and ventral attention systems. *Proc. Natl. Acad. Sci.* **103**, 10046–10051 (2006).

17. Vossel, S., Geng, J. J. & Fink, G. R. Dorsal and ventral attention systems: Distinct neural circuits but collaborative roles. *Neuroscientist* **20**, 150–159 (2014).

18. Roldán, É. & Parrondo, J. M. Entropy production and Kullback-Leibler divergence between stationary trajectories of discrete systems. *Phys. Rev. E* **85**, 031129 (2012).

19. Friston, K. J., Williams, S., Howard, R., Frackowiak, R. S. & Turner, R. Movement-related effects in fMRI time-series. *Magn. Reson. Med.* **35**, 346–355 (1996).

20. Power, J. D., Barnes, K. A., Snyder, A. Z., Schlaggar, B. L. & Petersen, S. E. Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. *Neuroimage* **59**, 2142–2154 (2012).

21. Prichard, D. & Theiler, J. Generating surrogate data for time series with several simultaneously measured variables. *Phys. Rev. Lett.* **73**, 951 (1994).

22. Barch, D. M. *et al.* Function in the human connectome: Task-fMRI and individual differences in behavior. *Neuroimage* **80**, 169–189 (2013).

23. Van Essen, D. C. *et al.* The WU-Minn Human Connectome Project: An overview. *Neuroimage* **80**, 62–79 (2013).

24. Schaefer, A. *et al.* Local-global parcellation of the human cerebral cortex from intrinsic functional connectivity MRI. *Cereb. Cortex* **28**, 3095–3114 (2018).

25. Siegel, J. S. *et al.* Data quality influences observed links between functional connectivity and behavior. *Cereb. Cortex* **27**, 4492–4502 (2017).