Selection entropy: The information hidden within neuronal patterns

Erik D. Fagerholm,[1,*] Zalina Dezhina,[1] Rosalyn J. Moran,[1] Karl J. Friston ⬤,[2] Federico Turkheimer ⬤,[2] and Robert Leech[1]

[1]*Department of Neuroimaging, King's College London, London, United Kingdom*
[2]*Wellcome Centre for Human Neuroimaging, London, United Kingdom*

This article is part of the Physical Review Research *collection titled Physics of Neuroscience.*

Boltzmann entropy is a measure of the hidden information contained within a system. In the context of neuroimaging, information can be hidden within the multiple brain states that cannot be distinguished within a single image. Here, we show that information can also be hidden within multiple indistinguishable selections of neuronal patterns between brain regions, as quantified by a novel metric that we term "selection entropy." We show the ways in which selection entropy behaves in comparison with the Kullback-Leibler (KL) divergence (relative entropy). First, we use synthetic data sets to demonstrate that selection entropy is more sensitive to small changes in probability distributions compared with the KL divergence. Second, we show that selection entropy identifies a principal gradient between sensorimotor and transmodal brain regions more definitively than the KL divergence within resting-state functional magnetic resonance imaging time series. As such, we introduce selection entropy as an additional asset in the analysis of neuronal functional selectivity.

I. INTRODUCTION

Self-organizing systems such as the brain can be described in terms of two principles—namely, functional segregation and integration [1]. Functional segregation refers to the specialization of different brain regions according to anatomically segregated cortical regions or neuronal populations. This specialization requires a selective expression of cortical activation repertoires over time.

This technical work examines the characterization of selective or distinct repertoires of activation in terms of information theoretic measures. In particular, we consider the standard approach based upon the Kullback-Leibler (KL) divergence (relative entropy) [2] between two probability distributions, and introduce a new measure that we call "selection entropy."

This paper comprises four sections. In the first section, we discuss the concept of Boltzmann entropy in the context of neuroimaging and show the ways in which it captures the hidden information contained within indistinguishable rearrangements of a system. In the second section, we introduce selection entropy—a metric that quantifies the hidden information hidden within the patterns selected *between* two systems. In the third section, we use synthetic data sets to show that selection entropy is more sensitive to small changes between probability distributions compared with the KL di-

vergence. In the fourth section, we use resting-state functional magnetic resonance imaging (fMRI) time series to show that selection entropy identifies a sensorimotor-to-transmodal gradient of cortical organization [3,4] more distinctly than the KL divergence.

A. Boltzmann entropy

Boltzmann entropy is a measure of hidden information [5] that is frequently used as a way of quantifying repertoires of neuronal states [6–26]. With regard to the concept of Boltzmann entropy, we consider a simple example of a neuroimaging modality that collects just two observations, each of which falls into one of two histogram bins, as shown in Fig. 1(a). We then ask the following question: in how many ways can we rearrange the observations in Fig. 1(a) so as to maintain the same distribution, i.e., one observation in each of the two bins? The answer is two, with the second possibility shown in Fig. 1(b).

In other words, if we only have access to the shape of the distribution (the "macrostate"), we are subject to an intrinsic ignorance that arises from the two indistinguishable ways (the "microstates") in which the observations can be rearranged. This ignorance is precisely what the Boltzmann entropy quantifies.

Let us now generalize the example in Fig. 1 such that there are a total of N observations. We then ask: in how many ways (W) can these N observations be rearranged to yield the same histogram shape, such that the first bin contains n_1 observations, the second bin contains n_2 observations, and so forth? The answer is given by the standard [27] combinatoric expression

$$W = \frac{N!}{\prod_i n_i!}, \tag{1}$$

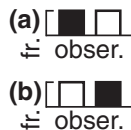*Corresponding author: erik.fagerholm@kcl.ac.uk

FIG. 1. (a) A histogram showing that observations (obser.) 1 and 2 (black and white) appear with frequency (fr.) 1 in the first and second bins, respectively. (b) The other rearrangement of (a) that yields the same distribution with one observation in each of the two bins.

where n_i is the number of observations (occupation number) of the ith bin.

For the example in Fig. 1, we can use Eq. (1) to see that there are indeed $W = \frac{2!}{1!1!} = 2$ ways of rearranging a histogram of $N = 2$ observations such that the first bin contains one observation and the second bin contains one observation.

Standard methods [28] can then be used to show that W in Eq. (1) leads directly to the following expression for the Boltzmann entropy, S, associated with a single observation:

$$S = -\sum_i p_i \log p_i, \qquad (2)$$

where $\log W \approx NS$, and $p_i = {}^{n_i}/_N$ is the probability of occurrence of the ith observation. Note that the expression in Eq. (2) only holds when N and all n_i are large.

See the Appendix for the details of the steps leading from Eq. (1) to Eq. (2).

B. Selection entropy

Let us now extend the toy model in Fig. 1 such that we deal with two histograms, each of which is populated by three observations that are distributed across two bins, as shown in Fig. 2(a).

We now ask an entirely different question to the one considered with regard to the Boltzmann entropy: in how many ways can the observations contained within the same bins in the two histograms be selected between one another? By inspecting Fig. 2(a), we see that, for each of the two ways in which the bin 1 observations in system 1 can be selected by system 2, there are two ways in which the bin 2 observations in system 2 can be selected by system 1 [Fig. 2(b)].

We can generalize the example in Fig. 2 by considering two histograms that contain N and M total observations. The total number of ways (W_{sel}) of selecting observations between like bins is given by the product of

(i) the number of ways of selecting the m_1 observations in the first bin of system M from the n_1 observations in the first bin of system N and

(ii) the number of ways of selecting the m_2 observations in the second bin of system M from the n_2 observations in the second bin of system N,

and so forth, as given by the product of the binomial coefficients

$$W_{sel} = \prod_i \frac{n_i!}{m_i!(n_i - m_i)!}, \qquad (3)$$

where n_i and m_i are the numbers of observations in the ith bins of the two histograms.

We can generalize the example in Fig. 2 by comparing two histograms constructed using data from a total of N_1 observations of system 1 and N_2 observations of system 2, respectively. Suppose that $n_i^{(1)}$ of the N_1 measurements of system 1 and $n_i^{(2)}$ of the N_2 measurements of system 2 fall into bin i. Let $n_i = \max(n_i^{(1)}, n_i^{(2)})$ be the larger of these two numbers and $m_i = \min(n_i^{(1)}, n_i^{(2)})$ be the smaller. The total number of ways of selecting m_i observations from n_i is $\frac{n_i!}{m_i!(n_i - m_i)!}$. The total number of ways (W_{sel}) of selecting the smaller number from the larger number of observations in all bins is given by the product of these binomial coefficients:

$$W_{sel} = \prod_i \frac{n_i!}{m_i!(n_i - m_i)!}. \qquad (4)$$

Using Eq. (3) for the example in Fig. 2(b), we see that there are indeed $\frac{2!}{1!(2-1)!} \times \frac{2!}{1!(2-1)!} = 4$ ways in which observations can be selected from like bins in the two histograms.

Following the same logic used to derive the Boltzmann entropy in Eq. (2) from the expression for rearrangements in Eq. (1) (see the Appendix), we begin by taking the logarithm of the expression for selections in Eq. (3):

$$\log W_{sel} = \sum_i (\log(n_i)! - \log(m_i)! - \log((n_i - m_i)!)). \qquad (5)$$

We then assume that the numbers of observations in all bins are sufficiently large to allow us to use Stirling's approximation [i.e., $\log(y!) \approx y \log y - y$], such that

$$\log W_{sel} \approx \sum_i \left(m_i \log\left(\frac{n_i}{m_i} - 1\right) - n_i \log\left(1 - \frac{m_i}{n_i}\right) \right). \qquad (6)$$

We can then express the numbers of observations within the ith bins of the two systems (m_i and n_i) in terms of the probabilities of finding the system in the ith bin [$p_M(m_i)$ and $p_N(n_i)$] and the total numbers of observations in the two systems (M and N) as follows: $n_i = p_N(n_i)N$ and
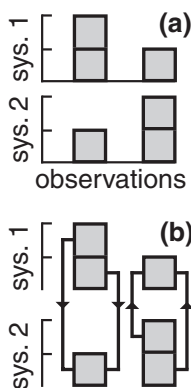

FIG. 2. (a) Two histograms for systems 1 and 2 showing that there are two observations in the first bin and one observation in the second bin for system (sys.) 1 and vice versa for system 2. (b) The same two histograms as in (a), showing that there are a total of four ways in which observations can be selected between like bins, as indicated by the arrows.

$m_i = p_M(m_i)M$, such that

$$\log W_{\text{sel}} \approx \sum_i \left(M p_M(m_i) \log\left(\frac{p_N(n_i)}{p_M(m_i)} \frac{N}{M} - 1 \right) \right.$$
$$\left. - N p_N(n_i) \log\left(1 - \frac{p_M(m_i)}{p_N(n_i)} \frac{M}{N} \right) \right). \quad (7)$$

We then define a ratio (r) of the observations,

$$r \equiv \frac{M}{N}, \quad (8)$$

which we use together with Eq. (7) to obtain

$$\log W_{\text{sel}} \approx N \sum_i \left(r p_M(m_i) \log\left(\frac{1}{r} \frac{p_N(n_i)}{p_M(m_i)} - 1 \right) \right.$$
$$\left. - p_N(n_i) \log\left(1 - r \frac{p_M(m_i)}{p_N(n_i)} \right) \right). \quad (9)$$

We then arrive at the following quantity, which we term the "selection entropy" associated with a single observation in the N length time series:

$$S_{\text{sel}} = \sum_i \left(r p_M(m_i) \log\left(\frac{1}{r} \frac{p_N(n_i)}{p_M(m_i)} - 1 \right) \right.$$
$$\left. - p_N(n_i) \log\left(1 - r \frac{p_M(m_i)}{p_N(n_i)} \right) \right). \quad (10)$$

It should be noted that, in the general case when $N \neq M$, the selection entropy depends on the numbers of observations made (N and M). In this sense, the selection entropy is not just a feature of the two probability distributions alone, as it depends on how many measurements are made of each system.

In summary, just as the Boltzmann entropy in Eq. (2) quantifies the hidden information contained within indistinguishable rearrangements of a single histogram via Eq. (2), the selection entropy in Eq. (10) quantifies the hidden information contained within indistinguishable selections between like bins of histogram pairs via Eq. (3).

C. KL divergence

The KL divergence quantifies the extent to which two probability distributions differ from one another. The KL divergence has frequently been used in the context of neuroscience [29,30] and, as such, together with the Boltzmann entropy, will form our main point of reference in characterizing selection entropy.

Using the same notation as in Eq. (10), the KL divergence from $p_M(m_i)$ to $p_N(n_i)$ is defined as follows:

$$KL = \sum_i p_N(n_i) \log\left(\frac{p_N(n_i)}{p_M(m_i)} \right). \quad (11)$$

When two distributions are identical, the KL divergence attains its lower bound of zero.

II. METHODS

A. Equal-length time series

We will henceforth calculate the selection entropy between pairs of equal-length neuroimaging time series. This means that the resultant histograms are constructed from equal numbers of data points, and hence from Eq. (8) $r \equiv \frac{M}{N} = 1$, which means that we use the following simplified version of Eq. (10):

$$S_{\text{sel}} = \sum_i \left(p_M(m_i) \log\left(\frac{p_N(n_i)}{p_M(m_i)} - 1 \right) \right.$$
$$\left. - p_N(n_i) \log\left(1 - \frac{p_M(m_i)}{p_N(n_i)} \right) \right). \quad (12)$$

B. Gaussian distributions

Equipped with this intuitive measure of hidden information, we can now use numerical analyses to understand better how selection entropy behaves in relation to the KL divergence. We will use samples from known distributions and empirical time series from resting-state functional magnetic resonance imaging. To create samples from known distributions we perform the following steps:

Step 1: We sample from a series of Gaussian distributions with mean zero and standard deviations that vary from one to two in small increments.

Step 2: We sample from a second set of Gaussian distributions that also have mean zero, but for which the standard deviations change in the opposite direction from two to one in small negative increments.

Step 3: We create histograms from the two sets of samples in steps 1 and 2, by entering each data point into one of 100 bins.

Step 4: We divide each of the histogram entries in step 3 by the total number of data points in each histogram to obtain (sample) probability distributions.

Step 5: We calculate the selection entropy and KL divergence for each pair of probability distributions in step 4.

C. Gradient-based resting-state fMRI

All the empirical time series were obtained from the 1200-subject release of the Human Connectome Project [31] in 30 young adults (age, 22 to 35 years). Images are acquired using a 3T Scanner (Siemens Skyra 3 Tesla MRI scanner; repetition time = 720 ms, time to echo = 33 ms, flip angle = 52°, voxel size = 2 mm isotropic, 72 slices, matrix = 104 × 90, field of view = 208 × 180 mm, multiband acceleration factor = 8).

We parcellated these data using the Schaefer 100 atlas [32] to obtain a resting-state functional connectivity matrix and then computed pairwise correlations between all cortical and subcortical time series to obtain normative functional connectivity matrices. These 100 regions were then reorganized according to the principal gradient identified by Margulies
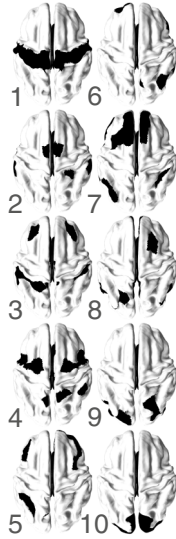
FIG. 3. The 100 regions of the atlas organized according to the sensorimotor-to- transmodal principal gradient. Images are shown in groups of 10, with image 1 displaying regions 1 through 10 in black, image 2 displaying regions 2 through 11, and so forth.

et al. [3,4], with sensorimotor regions at one end (region 1) and regions associated with higher cognitive function at the other end (Fig. 3).

D. fMRI data analysis

We compare the behavior of selection entropy and KL divergence for resting-state fMRI data via the following steps:

Step 1: We create histograms from each of the 100 time series for the 30 subjects by entering each data point into one of 100 bins.

Step 2: We divide each of the histogram entries in step 1 by the total number of data points in each histogram to obtain sample probability distributions.

Step 3: We calculate the selection entropy and KL divergence for all pairs of probability distributions in step 2, thus obtaining matrices of size $100 \times 100 \times 30$ (regions $\times$ regions $\times$ subjects).

Step 4: We smooth and average the matrices in step 3 across subjects to obtain two matrices of size 100×100 each, summarizing the selection entropy and KL divergence between all pairs of regions.

Step 5: We calculate the mean and standard errors of the first, second, etc., diagonals of the matrices in step 4 to obtain the change in selection entropy and KL divergence with increasing separation along the principal gradient.

III. RESULTS

All results described in what follows can be reproduced with the accompanying MATLAB code.

We first establish the behavior of selection entropy and KL divergence revealed by the synthetic time series (sampled from Gaussian distributions with varying standard deviations). We then review the corresponding results based upon the empirical (fMRI) time series.
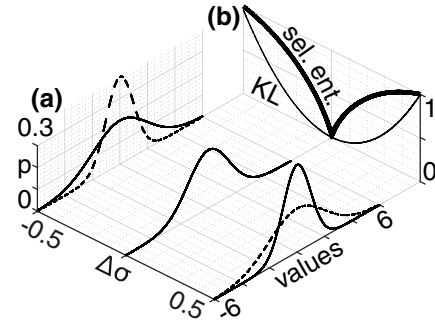

FIG. 4. (a) Two Gaussian curves with mean zero, with standard deviations that range between 1 and 2 (dashed line) and between 2 and 1 in small increments. The curves are shown at minimum, zero, and maximum differences in the standard deviations ($\Delta\sigma$). (b) Normalized values of the KL divergence (thin line) and the selection entropy (sel. ent.; thick line) for the same range of differences in the standard deviations shown in (a).

A. Gaussian distributions

As two Gaussian distributions become increasingly similar [Fig. 4(a)], we find that the selection entropy and KL divergence approach the origin in concave downward and concave upward curves, respectively [Fig. 4(b)].

Both the selection and KL divergence quantify the difference between the two distributions from which samples are taken. However, the selection entropy has an inflection point near the origin as the two distributions become arbitrarily close to one another. This indicates that the selection entropy is more sensitive to small differences between distributions.

B. Gradient-based resting-state fMRI

We base this section on probability distributions obtained from resting-state fMRI time series [Fig. 5(a)] collected in 100 regions organized according to a principal sensorimotor-to-transmodal gradient. We calculate the subject-averaged KL divergence and selection entropy for all pairs of these regions [Fig. 5(b)]. We then calculate the average KL divergence and selection entropy with increasing distance along the principal gradient [Fig. 5(c)].

We see from Fig. 5(c) that the selection entropy displays a more pronounced monotonic increase with increasing separation along the principal gradient, as compared with the KL divergence.

IV. DISCUSSION

We have shown that there are scenarios in which the selection entropy reveals different information to that provided by the commonly used KL divergence, as demonstrated by Figs. 4 and 5. Technically, we can characterize the relationship between selection entropy and KL divergence mathematically by expressing Eq. (12) as follows:

$$S_{\text{sel}} = \sum_i \left(p_N(n_i)\log\left(\frac{p_N(n_i)}{p_M(m_i)}\right) + (p_M(m_i) - p_N(n_i))\log\left(\frac{p_N(n_i)}{p_M(m_i)} - 1\right) \right), \quad (13)$$
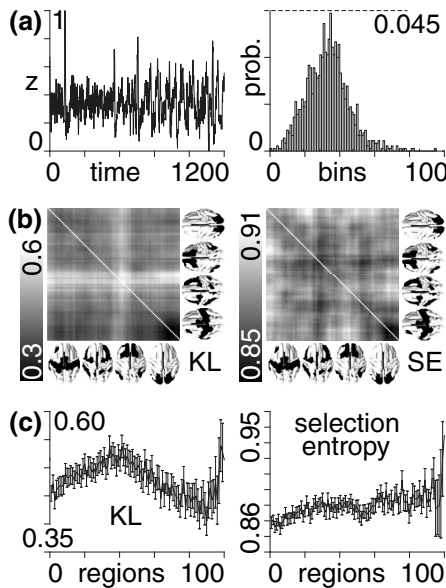
FIG. 5. (a) The time course for the first region of the first subject (left), with its associated probability distribution (right). (b) The smoothed KL divergence (left) and selection entropy (SE, right) for all 100×100 region pairs. The outer Brainspace images indicate quartiles of the principal gradient-ordered regions, i.e., the first image shows regions 1 through 25 in black, the second shows regions 26 through 50, etc. (c) Mean and standard error of the KL divergence (left) and selection entropy (right) for increasing separation between regions along the principal gradient.

where we recall from Eq. (11) that the first sum of terms are equal to the KL divergence, such that we can write the selection entropy as follows:

$$S_{\text{sel}} = KL + \sum_i (p_M(m_i) - p_N(n_i)) \log\left(\frac{p_N(n_i)}{p_M(m_i)} - 1\right);$$

(14)

i.e., the selection entropy can be viewed as a sum of terms that augment the KL divergence.

Similarly, we can characterize the relationship between Boltzmann entropy and selection entropy mathematically by rewriting Eq. (12) as follows:

$$S_{\text{sel}} = \sum_i p_N(n_i) \log p_N(n_i) - \sum_i p_M(m_i) \log p_M(m_i)$$
$$- \sum_i (p_N(n_i) - p_M(m_i)) \log(p_N(n_i) - p_M(m_i)),$$

(15)

which comprises the following three terms: (i) the Boltzmann entropy for system N,

$$S_N = -\sum_i p_N(n_i) \log p_N(n_i);$$

(16)

(ii) the Boltzmann entropy for system M,

$$S_M = -\sum_i p_M(m_i) \log p_M(m_i);$$

(17)

and (iii) the Boltzmann entropy for the difference between systems N and M,

$$S_{N-M} = -\sum_i (p_N(n_i) - p_M(m_i)) \log(p_N(n_i) - p_M(m_i)).$$

(18)

Using Eqs. (15) through (18), e can therefore write the selection entropy as the following linear combination of Boltzmann entropies:

$$S_{\text{sel}} = S_M - S_N + S_{N-M}.$$

(19)

This shows that, in addition to the hidden information in the two systems ($S_M - S_N$), the selection entropy also contains an interaction term (S_{N-M}) that depends upon the difference in probabilities between the two systems.

The relationship between Boltzmann entropy and selection entropy can be expanded upon in terms of the disparity between measurements and indistinguishable configurations. For instance, the Boltzmann entropy quantifies the hidden information associated with an ambiguity between a macroscopic measurement ("macrostate") and multiple indistinguishable microscopic rearrangements ("microstates") of the underlying system. We can view this ambiguity in terms of the multiple ways of shuffling the constituent masses of a probability distribution while keeping its shape unchanged [see Fig. (1)]. On the other hand, the selection entropy quantifies the hidden information associated with an ambiguity between what we can call a "macroselection" (i.e., the knowledge that the states of one system are selected from another) and what we can call "microselections" (i.e., the multiple ways in which like state components of two systems can be cross-selected) (see Fig. 2).

As opposed to the KL divergence, the selection entropy displays an inflection point as probability distributions become very similar, as shown with the synthetic data (Fig. 4). This suggests that the selection entropy is more sensitive to small changes in distributions. Furthermore, the selection entropy is more sensitive to changes between time series-derived probability distributions collected from neuronal regions relative to their location along a principal gradient of cortical organization [Fig. 5(c)]. As such, we suggest that selection entropy could be useful in the study of functional segregation, due to its ability to reveal novel information within neuroimaging time series.

All data and accompanying MATLAB code are available at the following public repository [33].

APPENDIX

We begin by taking the logarithm of W in Eq. (1):

$$\log W = \log\left(\frac{N!}{\prod_{i=1}^{x} n_i!}\right) = \log(N!) - \log\left(\prod_{i=1}^{x} n_i!\right)$$

$$= \log(N!) - \sum_{i=1}^{x} \log(n_i!). \tag{A1}$$

We then assume that N and all n_i are large, such that we can use Stirling's approximation $[\log(y!) \approx y\log y - y]$ to give

$$\log W \approx N\log N - N - \sum_{i=1}^{x}(n_i\log n_i - n_i)$$

$$= N\log N - N - \sum_{i=1}^{x} n_i\log n_i + \sum_{i=1}^{x} n_i, \tag{A2}$$

where $\sum_i n_i = N$, such that

$$\log W \approx N\log N - N - \sum_{i=1}^{x} n_i\log n_i + N$$

$$= N\log N - \sum_{i=1}^{x} n_i\log n_i. \tag{A3}$$

We can now express the number of observations occupying the ith bin (n_i) in terms of the probability of finding an observation in the ith bin (p_i) as $n_i = Np_i$, such that

$$\log W \approx N\log N - \sum_{i=1}^{x} Np_i\log(Np_i)$$

$$= N\log N - \sum_{i=1}^{x} Np_i(\log N + \log p_i)$$

$$= N\log N - \sum_{i=1}^{x} Np_i\log N - \sum_{i=1}^{x} Np_i\log p_i$$

$$= N\log N - N\log N \sum_{i=1}^{x} p_i - N\sum_{i=1}^{x} p_i\log p_i, \tag{A4}$$

where $\sum_i p_i = 1$, such that

$$\log W \approx N\log N - N\log N - N\sum_{i=1}^{x} p_i\log p_i$$

$$= -N\sum_{i=1}^{x} p_i\log p_i; \tag{A5}$$

i.e., the logarithm of the number of ways W of rearranging the observations for a given set of occupation numbers n_i is directly proportional to the entropy associated with a single observation $-\sum_{i=1}^{x} p_i\log p_i$.

[1] S. Zeki and S. Shipp, The functional logic of cortical connections, Nature (London) **335**, 311 (1988).

[2] I. Csiszár, I-divergence geometry of probability distributions and minimization problems, Ann. Probab. **3**, 146 (1975).

[3] D. S. Margulies, S. S. Ghosh, A. Goulas, M. Falkiewicz, J. M. Huntenburg, G. Langs, G. Bezgin, S. B. Eickhoff, F. X. Castellanos, M. Petrides, E. Jefferies, and J. Smallwood, Situating the default-mode network along a principal gradient of macroscale cortical organization, Proc. Natl. Acad. Sci. **113**, 12574 (2016).

[4] J. M. Huntenburg, P.-L. Bazin, and D. S. Margulies, Large-scale gradients in human cortical organization, Trends Cogn. Sci. **22**, 21 (2018).

[5] H. C. Van Ness, *Understanding Thermodynamics* (Courier Corporation, Dover, New York, 1983).

[6] S. Keshmiri, Entropy and the brain: An overview, Entropy **22**, 917 (2020).

[7] D. J. Wang, K. Jann, C. Fan, Y. Qiao, Y.- F. Zang, H. Lu, and Y. Yang, Neurophysiological basis of multi-scale entropy of brain complexity and its relationship with functional connectivity, Front. Neurosci. **12**, 352 (2018).

[8] H. Yang, W. L. Shew, R. Roy, and D. Plenz, Maximal variability of phase synchrony in cortical networks with neuronal avalanches, J. Neurosci. **32**, 1061 (2012).

[9] D. P. Shorten, R. E. Spinney, and J. T. Lizier, Estimating transfer entropy in continuous time between neural spike trains

or other event-based data, PLoS Comput. Biol. **17**, e1008054 (2021).

[10] Y. Song and J. Zhang, Discriminating preictal and interictal brain states in intracranial EEG by sample entropy and extreme learning machine, J. Neurosci. Methods **257**, 45 (2016).

[11] R. E. Spinney, M. Prokopenko, and J. T. Lizier, Transfer entropy in continuous time, with applications to jump and neural spiking processes, Phys. Rev. E **95**, 032319 (2017).

[12] S. P. Strong, R. Koberle, R. R. de Ruyter van Stevenick, and W. Bialek, Entropy and Information in Neural Spike Trains, Phys. Rev. Lett. **80**, 197 (1998).

[13] J. B. Thomas, M. R. Brier, M. Ortega, T. L. Benzinger, and B. M. Ances, Weighted brain networks in disease: centrality and entropy in human immunodeficiency virus and aging, Neurobiol. Aging **36**, 401 (2015).

[14] M. Ursino, G. Ricci, and E. Magosso, Transfer entropy as a measure of brain connectivity: A critical analysis with the help of neural mass models, Front. Comput. Neurosci. **14**, 45 (2020).

[15] A. Viol, F. Palhano-Fontes, H. Onias, D. B. de Araujo, and G. Viswanathan, Shannon entropy of brain functional complex networks under the influence of the psychedelic Ayahuasca, Sci. Rep. **7**, 7388 (2017).

[16] R. M. Vivot, C. Pallavicini, F. Zamberlan, D. Vigo, and E. Tagliazucchi, Meditation increases the entropy of brain oscillatory activity, Neuroscience **431**, 40 (2020).

[17] T. Wadhera and D. Kakkar, Conditional entropy approach to analyze cognitive dynamics in autism spectrum disorder, Neurol. Res. **42**, 869 (2020).

[18] B. Wang, Y. Niu, L. Miao, R. Cao, P. Yan, H. Guo, D. Li, Y. Guo, T. Yan, J. Wu, J. Xiang, and H. Zhang Decreased complexity in Alzheimer's disease: resting-state fMRI evidence of brain entropy mapping, Front. Aging Neurosci. **9**, 378 (2017).

[19] Z. Wang, Y. Li, A. R. Childress, and J. A. Detre, Brain entropy mapping using fMRI, PLoS One **9**, e89948 (2014).

[20] T. Watanabe, S. Hirose, H. Wada, Y. Imai, T. Machida, I. Shirouzu, S. Konishi, Y. Miyashita, and N. Masuda, A pairwise maximum entropy model accurately describes resting-state human brain networks, Nat. Commun. **4**, 1370 (2013).

[21] M. Wibral, R. Vicente, and M. Lindner, Transfer Entropy in Neuroscienc in *Directed Information Measures in Neuroscience*, edited by M. Wibral, R. Vicente, J. T. Lizier (Springer Berlin, Heidelberg, Germany, 2014), p. 3–32.

[22] P. Wollstadt, M. Martínez-Zarzuela, R. Vicente, F. J. Díaz-Pernas, and M. Wibral, Efficient transfer entropy analysis of non-stationary neural time series, PLoS One **9**, e102833 (2014).

[23] Y. Wu, P. Chen, X. Luo, M. Wu, L. Liao, S. Yang, and R. M. Rangayyan, Measuring signal fluctuations in gait rhythm time series of patients with Parkinson's disease using entropy parameters, Biomed. Signal Process. Control **31**, 265 (2017).

[24] Y. Yao, W. Lu, B. Xu, C. Li, C. Lin, D. Waxman, and J. Feng, The increase of the functional entropy of the human brain with age, Sci. Rep. **3**, 2853 (2013).

[25] H. Yuan, F. Xiong, and X. Huai, A method for estimating the number of hidden neurons in feed-forward neural networks based on information entropy, Comput. Electron. Agric. **40**, 57 (2003).

[26] F. Zhou, Y. Zhuang, H. Gong, J. Zhan, M. Grossman, and Z. Wang, Resting state brain entropy alterations in relapsing remitting multiple sclerosis, PLoS One **11**, e0146080 (2016).

[27] K. P. Bogart, *Combinatorics through Guided Discovery*, (CreateSpace Independent Publishing Platform, Scotts Valley, California, US, 2017).

[28] R. C. Tolman, *The Principles of Statistical Mechanics* (Courier Corporation, North Chelmsford, Massachusetts, 1979).

[29] M. Wang, J. Jiang, Z. Yan, I. Alberts, J. Ge, H. Zhang, C. Zuo, J. Yu, A. Rominger, and K. Shi, Individual brain metabolic connectome indicator based on Kullback-Leibler divergence similarity estimation predicts progression from mild cognitive impairment to Alzheimer's dementia, Eur. J. Nucl. Med. Mol. Imaging **47**, 2753 (2020).

[30] L. Yin, G. Zhang, and F. Yin, Measuring synchronization between spikes and local field potential based on the Kullback–Leibler divergence, Comput. Intell. Neurosci. **2021**, 9954302 (2021).

[31] D. C. Van Essen, S. M. Smith, D. M. Barch, T. E. J. Behrens, E. Yacoub, and K. Ugurbil, for the WU-Minn HCP Consortium, The WU-Minn human connectome project: An overview, Neuroimage **80**, 62 (2013).

[32] A. Schaefer, R. Kong, E. M. Gordon, T. O. Laumann, X.-N. Zuo, A. J. Holmes, S. B. Eickhoff, and B. T. T. Yeo, Local-global parcellation of the human cerebral cortex from intrinsic functional connectivity MRI, Cereb. Cortex **28**, 3095 (2018).

[33] https://github.com/allavailablepubliccode/Selection_Entropy.