



Quantifying Dynamical High-Order Interdependencies From the O-Information: An Application to Neural Spiking Dynamics

Sebastiano Stramaglia^{1,2*}, Tomas Scagliarini¹, Bryan C. Daniels³ and Daniele Marinazzo⁴

¹ Dipartimento Interateneo di Fisica, Università degli Studi Aldo Moro, Bari and INFN, Bari, Italy, ² Center of Innovative Technologies for Signal Detection and Processing (TIRES), Università degli Studi Aldo Moro, Bari, Italy, ³ Arizona State University and Santa Fe Institute Center for Biosocial Complex Systems, Arizona State University, Tempe, AZ, United States, ⁴ Department of Data Analysis, Ghent University, Ghent, Belgium

OPEN ACCESS

Edited by:

Plamen Ch. Ivanov,
Boston University, United States

Reviewed by:

Paolo Grigolini,
University of North Texas,
United States

Fernando Rosas,
Imperial College London,
United Kingdom

*Correspondence:

Sebastiano Stramaglia
sebastiano.stramaglia@ba.infn.it

Specialty section:

This article was submitted to
Fractal and Network Physiology,
a section of the journal
Frontiers in Physiology

Received: 17 August 2020

Accepted: 16 December 2020

Published: 14 January 2021

Citation:

Stramaglia S, Scagliarini T, Daniels BC
and Marinazzo D (2021) Quantifying
Dynamical High-Order
Interdependencies From the
O-Information: An Application to
Neural Spiking Dynamics.
Front. Physiol. 11:595736.
doi: 10.3389/fphys.2020.595736

We address the problem of efficiently and informatively quantifying how multiplets of variables carry information about the future of the dynamical system they belong to. In particular we want to identify groups of variables carrying redundant or synergistic information, and track how the size and the composition of these multiplets changes as the collective behavior of the system evolves. In order to afford a parsimonious expansion of shared information, and at the same time control for lagged interactions and common effect, we develop a dynamical, conditioned version of the O-information, a framework recently proposed to quantify high-order interdependencies via multivariate extension of the mutual information. The dynamic O-information, here introduced, allows to separate multiplets of variables which influence synergistically the future of the system from redundant multiplets. We apply this framework to a dataset of spiking neurons from a monkey performing a perceptual discrimination task. The method identifies synergistic multiplets that include neurons previously categorized as containing little relevant information individually.

Keywords: information theory, transfer entropy, dynamical systems, spiking neurons, partial information decomposition

1. INTRODUCTION

High-order interdependencies are at the core of complex systems. In many biological systems, pairwise interactions have been found to be insufficient for explaining the orchestrated activity of multiple components (Crutchfield, 1994; Ohiorhenuan et al., 2010; Katz et al., 2011; Yu et al., 2011; Daniels et al., 2016). This may be crucial also in relation with the important question of how biological systems dynamically interact and collectively behave as a network to produce health or disease, the core business of network physiology (Bashan et al., 2012; Lin et al., 2020). The abundance of available data is pushing nowadays the development of effective algorithms for the inference of higher order interactions from data (Bettencourt et al., 2008; Stramaglia et al., 2012). When an information theoretical point of view is adopted, the problem of higher order interactions becomes related with the decomposition of the information flow in redundant and synergistic components, an issue which cannot be addressed within the Shannon framework unless further assumptions are made (Williams and Beer, 2010). Partial Information Decomposition (PID)

algorithms have been proposed (Bertschinger et al., 2014; Barrett, 2015; Lizier et al., 2018) based on the idea that synergies are statistical relationships which can be seen only if the whole set of driving variables is considered. Unfortunately, the practical use of PID is greatly limited by the super-exponential growth of terms for large systems, and many works limit the analysis to triplets of variables (Marinazzo et al., 2019).

All the approaches described above assume that polyadic relationships are important for the complex system under consideration and they thus query the validity of complex networks (taking into account only dyadic relationships) as a theoretical model for large-scale systems, James et al. (2016). For the estimate of information flow, the dyadic quantity which is commonly used is the transfer entropy (Schreiber, 2000), related to the concept of Granger causality (Barnett et al., 2009), which has been proposed to distinguish effectively driving and responding elements and to detect asymmetry in the interaction of subsystems. With the appropriate conditioning of transition probabilities this quantity has been shown to perform better than time delayed mutual information to infer interactions, as delayed correlations often fail to distinguish information that is actually exchanged from shared information due to common history and input signals (Bossomaier et al., 2016). Attempts to generalize the notion of transfer entropy beyond the network description have been made: the expansion of the transfer entropy from a multiplet of variables to a given target has been developed in Stramaglia et al. (2012) to highlight subgroups of variables which provide redundant and synergistic information to the target. For triplets of variables, the exact calculation of multiscale PID for Gaussian processes has been presented in Faes (2017). Merging concepts of PID and integrated information, the integrated decomposition framework has been developed in Mediano et al. (2019).

In a recent paper (Rosas et al., 2019), a novel quantity has been introduced to study statistical synergy, the O-information, a metric capable of characterizing synergy- and redundancy-dominated systems and whose computational complexity scales gracefully with system size, making it suitable for practical data analysis; the O-information has been used to study brain aging in Gatica et al. (2020). We remark that the O-information uses equal-time samples of variables, so its output depends only on equal-time correlations in the data-set and is insensitive to dynamic transfer of information; moreover the estimation of O-information does not require a division between predictor and target variables.

In this work we propose a dynamical generalization of the O-information to handle multivariate time series which, apart from equal-time correlations, takes into account also the lagged correlations with a given variable which is assumed to be the target. The proposed approach highlights informational circuits which dynamically influence the target variable in a synergistic or redundant fashion, with a much lighter computational burden, for large systems, than those required by the exact expansion of Stramaglia et al. (2012) or PID approaches in the spirit of Williams and Beer (2010) or the dynamic frame PhiID introduced in Mediano et al. (2019). We apply this quantity, that will be denoted as dO-information, to study the neural spiking dynamics recorded from a multielectrode array with 169

channels during a visual motion direction discrimination task, which has been already considered in Daniels et al. (2017) in the frame of Dual Coding Theory; here will analyze this data-set with the aim of characterizing how the dynamic transfer of information is shaped by redundant and synergistic multiplets of variables. Our main result is that a class of neurons, not encoding any information on the decision at the individual level, are otherwise important in the construction of synergistic circuits.

2. METHODS

2.1. Dynamic O-Information

Given a collection of n random variables $\mathbf{X} = \{X_1, \dots, X_n\}$, the O-information (shorthand for “information about Organizational structure”) is defined as follows (Rosas et al., 2019):

$$\Omega_n = (n - 2)H(\mathbf{X}) + \sum_{j=1}^n [H(X_j) - H(\mathbf{X} \setminus X_j)], \quad (1)$$

where H stands for the entropy, and $\mathbf{X} \setminus X_j$ is the complement of X_j with respect to \mathbf{X} . If $\Omega_n > 0$ the system is redundancy-dominated, while if $\Omega_n < 0$ it is synergy-dominated. Let us now add the stochastic variable Y to the set of \mathbf{X} variables. The O-information now reads

$$\Omega_{n+1} = \Omega_n + \Delta_n,$$

where

$$\Delta_n = (1 - n)I(Y; \mathbf{X}) + \sum_{j=1}^n I(Y; \mathbf{X} \setminus X_j), \quad (2)$$

I denoting the mutual information. Δ_n is the variation of the total O-information when the new variable Y is added, measuring the informational character of the circuits which link Y with variables \mathbf{X} : if Δ_n is positive, then y receives mostly redundant information from \mathbf{X} variables, whilst a negative Δ_n means that the influence of \mathbf{X} on y is dominated by synergistic effects.

Let us now consider a multivariate set of n time series $\{x_k\}_{k=1, \dots, n}$ and a target series y . Choosing an order m for the time series, we consider as the random variables \mathbf{X} the state vectors

$$X_k(t) = (x_k(t) \ x_k(t-1) \ \dots \ x_k(t-m+1));$$

where varying the time index t we get different samples of \mathbf{X} . The role of the variable Y is now played by the target time series, i.e., samples of Y are obtained as $Y(t) = y(t+1)$ varying the time index t . With these definitions, Δ_n measures the character of the information flow from the x variables to the target y . However, in order to remove shared information due to common history and input signals, one should condition on the state vector of the target variable, thus leading to the definition of the dynamic O-information from the group of variables $\{x_k\}_{k=1, \dots, n}$ to the target series y :

$$d\Omega_n = (1 - n)I(Y; \mathbf{X}|Y_0) + \sum_{j=1}^n I(Y; \mathbf{X} \setminus X_j|Y_0), \quad (3)$$

where $Y_0(t) = (y(t) y(t-1) \dots y(t-m+1))$ are the samples of Y_0 .

Some properties of the dO-information are in order. In the case of two driving variables, we have:

$$d\Omega_2 = I(y; X_1|Y) + I(y; X_2|Y) - I(y; X_1X_2|Y), \quad (4)$$

coinciding with the second order term of the expansion of the transfer entropy developed in Stramaglia et al. (2012). Expression (4) may be seen as a dynamical generalization of the interaction information, a well-known information measure for sets of three variables (McGill, 1954). Another property: let us suppose that the variable x_n is statistically independent of the others, i.e.,

$$p(y, Y, X_1, X_2, \dots, X_n) = p(y, Y, X_1, X_2, \dots, X_{n-1}) p(X_n), \quad (5)$$

then the dynamic O-information does not change under inclusion of x_n , i.e.,

$$d\Omega_n = d\Omega_{n-1}. \quad (6)$$

Since $d\Omega_1 = 0$, the property above ensures that the dO-information is not sensitive to pure pairwise interactions.

2.2. The Optimization Problem

We use the expression of the dO-information to define the optimization problem of determining the set of k variables which maximizes $d\Omega_k$, with $k < n$; this search leads to the most redundant circuit of $k+1$ variables, assuming y as the target. The value of the dO-information for the multiplet of k variables maximizing $d\Omega_k$ will be called *redundancy*.

Analogously, the search for the set of k variables which minimizes $d\Omega_k$ leads to the most synergistic circuit of $k+1$ variables. The opposite of value of the dO-information for the multiplet of k variables minimizing $d\Omega_k$ will be called *synergy*.

We remark that what we call *redundancy* and *synergy* here refers to the dynamic (conditioned on the past) version of the quantities defined in Rosas et al. (2019), and should thus contain a *d-* prefix. For simplicity and in the spirit of what already accepted in the Partial Information Decomposition field, we decided to omit it.

As the extensive search for these motifs is unfeasible for large k , we adopt a greedy search strategy, where the extensive search is performed for $k = 2$, and larger k are handled adding one variable at a time to the best multiplet of $k-1$ variables.

In order to define a criterion to stop the greedy search for the redundant k variables motifs, one can estimate the probability that the increment $d\Omega_k - d\Omega_{k-1}$ is lower than those corresponding to the inclusion of a randomized time series (obtained, e.g., by a random circular shift of the k -th selected x time series). The k -th variables is thus added to the multiplet when such probability is lower than a given threshold. A similar criterion can be applied also to stop the search for synergistic k variables motifs.

2.3. Toy Example

As a toy example, let consider a system of four binary variables $\sigma_i(t)$ such that σ_1 σ_2 and σ_3 are 0 or 1 with equal probability at

each time, whilst $P(\sigma_4(t+1)|\sigma_1(t), \sigma_2(t), \sigma_3(t))$ is given by the following probabilities:

$$\begin{aligned} P(0|0, 0, 0) &= 1 - a + b, \\ P(0|0, 1, 0) &= a - b, \\ P(0|1, 1, 0) &= 1 - a + b, \\ P(1|1, 1, 0) &= a - b, \\ P(1|0, 1, 0) &= 1 - a + b, \\ P(0|1, 0, 0) &= a - b, \\ P(1|0, 0, 0) &= a - b, \\ P(1|1, 0, 0) &= 1 - a + b, \\ P(0|0, 0, 1) &= a + b, \\ P(0|0, 1, 1) &= 1 - a - b, \\ P(0|1, 1, 1) &= a + b, \\ P(1|1, 1, 1) &= 1 - a - b, \\ P(1|0, 1, 1) &= a + b, \\ P(0|1, 0, 1) &= 1 - a - b, \\ P(1|0, 0, 1) &= 1 - a - b, \\ P(1|1, 0, 1) &= a + b. \end{aligned}$$

The conditional probabilities reported above have been chosen so as to have two informational circuits in the system: the variable σ_4 receives dynamically synergistic information, by construction, from the pair $\{\sigma_1, \sigma_2\}$ and from the triplet $\{\sigma_1, \sigma_2, \sigma_3\}$, depending, respectively, on the parameters b and a .

Indeed, for $b = 0$, with probability a the variable σ_4 at time $t+1$ is given by the majority rule applied to the three driving variables at time t , unless the three variables are equal: if the three driving variables are all equal, then σ_4 becomes the opposite of them with probability a . Therefore the information provided by $\{\sigma_1, \sigma_2, \sigma_3\}$ is synergistic and all the three variables must be known in order to improve the predictability of σ_4 .

On the other hand, for $a = 0$, σ_4 , with probability b , is given by the XOR applied to $\{\sigma_1, \sigma_2\}$. When both a and b are non vanishing, two synergistic circuits of three and two variables influence the target σ_4 .

In **Figure 1**, we depict the $d\Omega_3$ from the triplet as well as the $d\Omega_2$ from the pair of synergistic drivers, for $a = 0.7$ and varying b . For $b = 0$, just the triplet $\{\sigma_1, \sigma_2, \sigma_3\}$ is correctly recognized as driving the target. As b increases, also the pair $\{\sigma_1, \sigma_2\}$ is recognized as a synergistic driver. Note that also $d\Omega_3$ decreases with b : indeed the dynamic O-information $d\Omega_n$ by construction sums up the contributions from the informational circuits corresponding to subsets of the n variables.

Crucially, in situations like this one, where the information flow is dynamic, the O-information fails to provide a description of the system, and a dynamical approach like $d\Omega$ is mandatory.

2.4. Further Comments on Methods

The dO-information is designed to probe higher-order lagged influences, and provides a picture of a complex system complementary to those provided by the information flow

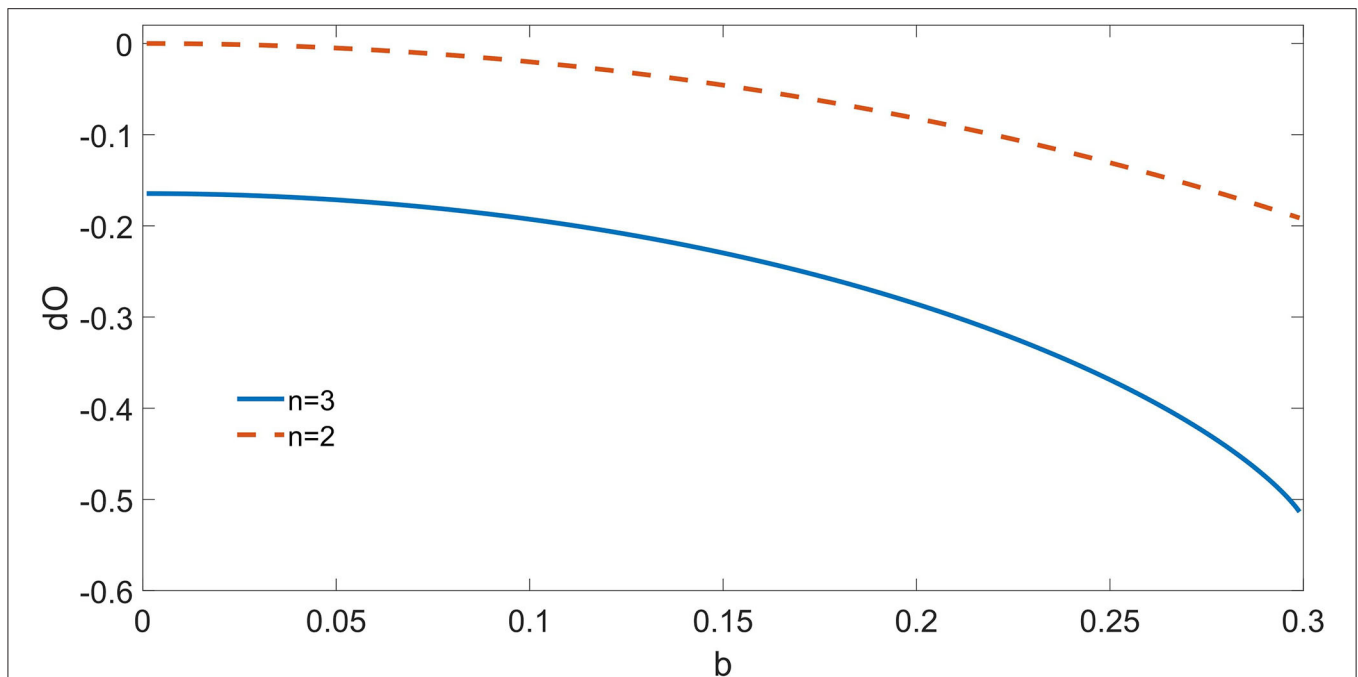


FIGURE 1 | For the toy model described in the text, consisting of four binary variables, we depict the dynamic O-Information from the pair of drivers and from the triplet of synergistic drivers as a function of the parameter b ; a is fixed at 0.7. At $b = 0$, only the multiplet of three drivers is recognized by the proposed approach; as b increases, also the circuit of two drivers σ_1 - σ_2 is recognized. The proposed method is thus capable to reveal the simultaneous presence of two informational circuits influencing the target variable.

network as measured by the pairwise transfer entropy (Schreiber, 2000):

$$TE(x_i \rightarrow z) = I(y; X_i|Y).$$

It is worth mentioning that the relation between dO-information and transfer entropy is of the same nature as the one between O-information and mutual information (the *condition-on-the-past* operator is applied to both in the same way). In order to assess the significance of the pairwise transfer entropy, for each pair driver-target we consider surrogate interactions (obtained by blockwise circular shift of the target time series) and accept a non-zero value only if the probability that randomization of the target leads to a value of the transfer entropy higher than the measured one is < 0.05 . Recently another estimator, based on a theoretical framework for TE in continuous time and extended to event based data, connected to a local permutation surrogate generation strategy, was proposed (Shorten et al., 2020).

For the application considered in this work, we have computed the conditioned mutual information terms, composing the dO-informations, using the Gaussian Copulas approach described in Ince et al. (2017).

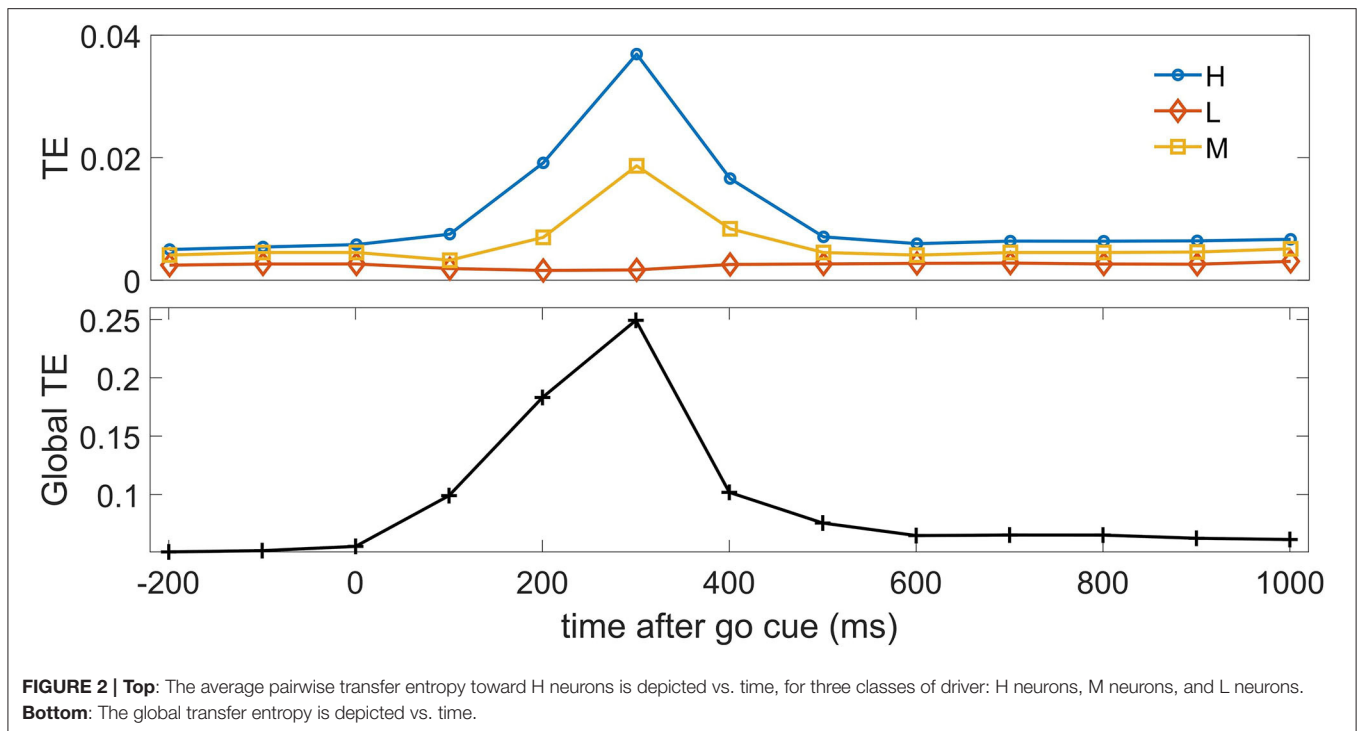
It is worth stressing the conceptual difference between the search for the most informative variables for a given target (i.e., the n variables \mathbf{X} maximizing $I(y; \mathbf{X}|Y)$), and the search for the most synergistic multiplet (i.e., the variables \mathbf{X} minimizing $d\Omega_n$). Suppose that, during the greedy search, one has already selected $n-1$ variables and now has to look for the n -th variable. If the new variable is selected so as to maximize the information about

the target, then the information gain due to its inclusion may be due to synergistic interactions with the previously selected $n-1$ variables, or to unique information from the new variable, where unique information means a contribution to the predictability of the target that can be obtained from that variable even when it is treated as the only driver. Inclusion of variables providing unique information does not give us further insights about the system beyond what we already knew from the pairwise description. Minimization of $d\Omega_n$ is instead tailored to take into account only synergistic interactions, and thus to elicit informational circuits of variables which influence the target: in other words, minimizing $d\Omega_n$ leads to discover (even small) improvements of the predictability of the target which can be ascribed to the joint action of groups of driving variables, thus allowing a picture of the system beyond the pairwise description.

We like also to stress that our approach may be also seen as a computationally light method for an exploratory search of synergistic multiplets of variables, and in this sense it might also be seen as a pre-processing stage for further processing of the multiplet, e.g., by the approach PhiID described in Mediano et al. (2019) to address issues like dynamical complexity or integrated information.

3. DATASET

We use data from the Random Dot Motion discrimination task (Shadlen and Newsome, 2001; Kiani and Shadlen, 2009; Kiani et al., 2014, 2015), in which the subject must decide which



direction dots on a screen are moving. This dataset comes from the sample T33 on the data sharing website <https://www.cns.nyu.edu/kianilab/Datasets.html> and has been already analyzed in an information theory framework in Daniels et al. (2017). We analyze the activity of 169 neural channels in a macaque monkey performing the task, across 1,778 trials. Data consist of spike times measured at a resolution of 1/30 ms, subsequently binned in intervals spaced 100 ms. In each trial, after the perceptual stimulus, a go cue is given to the subject to prompt it to indicate its decision. In Daniels et al. (2017) the analysis had decision as the target, and neurons were divided in three groups according to the information that their dynamics provide about the decision: those in Class H encode information before the go cue, those in Class M encode information after but not before the go cue, and those in Class L never encode information. Here, we will not take into account the decision as a variable, retaining only the classification of neurons in the three classes H, M, and L.

As an example application of the proposed method, we will consider the internal dynamics of the neuronal system. For each H neuron, taken as the target, we will study the higher order interactions from the rest of the measured neurons, concentrating on the most redundant circuits as well as the most synergistic ones.

4. RESULTS

For the following analyses we used $m = 1$ as lag for conditioning in the past, namely a bin one time step in the past, because the temporal correlations in data rapidly decay

at lag 2, corresponding to 200 ms.. In Figure 2, we depict the pairwise transfer entropy as a function of time, where the target is an H neuron and the driver is a neuron belonging to one of the three classes; the curves are averaged over the target neuron and over the driver belonging to each of the three classes. We observe that the information flow peaks around 300 ms after the go cue, when the saccade has on average just happened, and the information on the decision in the neurons H and M is maximized, and that most of the effective influence arise from the other H neurons and (to a lesser extent) from M neurons. The pairwise transfer entropy from L neurons is negligible, hence at the bivariate level L neurons seem to play no role in the construction of the dynamical response of the system: indeed by definition L neurons, differently from H and M neurons, do not carry information about the decision. The lower panel of the figure depicts the global transfer entropy (Barnett et al., 2013) averaged over all the H neurons as targets; the global transfer entropy measures the information flow about the target when all the other variables are simultaneously taken as the driving set.

Let us now turn to consider higher order interactions, and start with the O-information, the approach introduced in Rosas et al. (2019) which considers only equal time correlations. In Figure 3, we depict, as a function of time, the O-information of the three sets of neurons as well as the O-information of the whole system of neurons. We note that H neurons (and M neurons as well) increase their redundancy (as measured by O-information) with a latency of 400 ms, where also the whole system of neurons displays a clear peak. On the other hand the system of L neurons do not show any reaction to the go cue at the level of O-information. Hence Figure 3 shows that equal

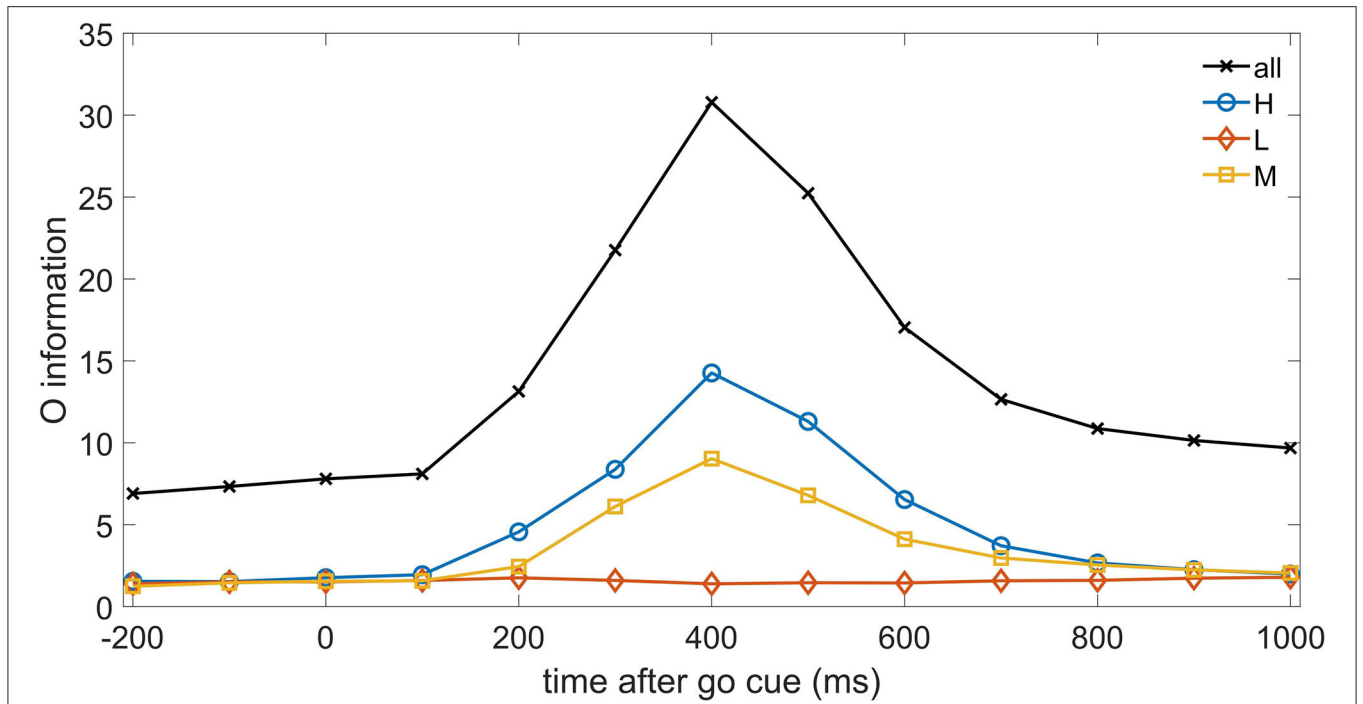


FIGURE 3 | The O-information is depicted vs. time for the three systems of neurons H, M, and L, as well as for the whole system of neurons.

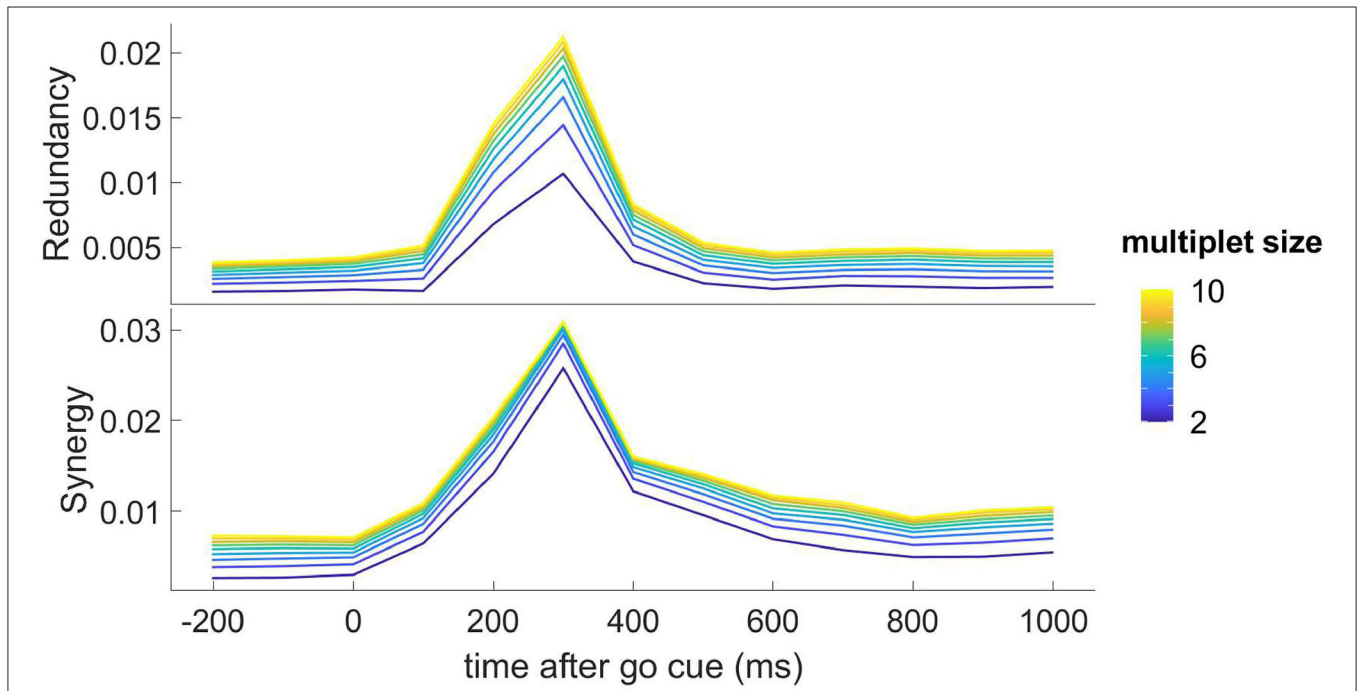
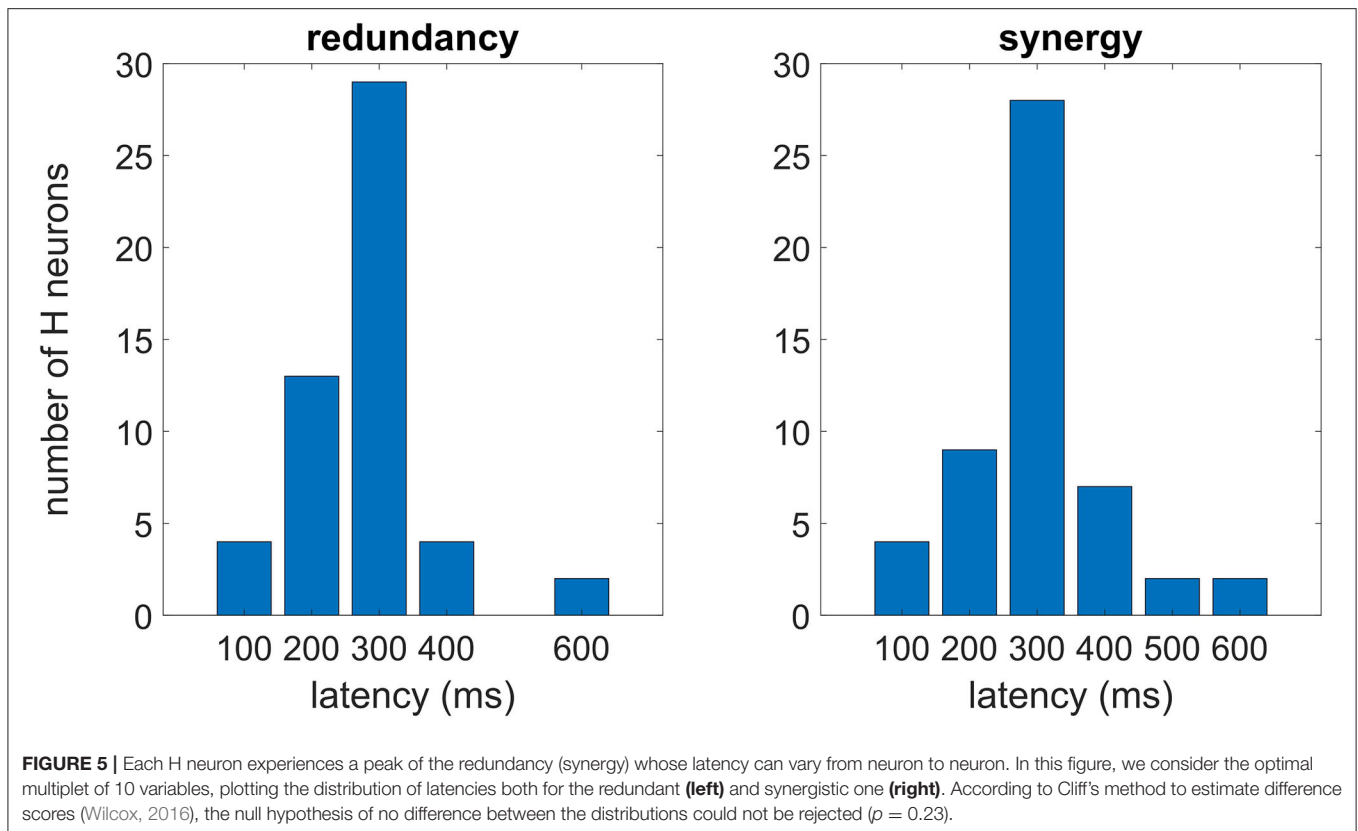


FIGURE 4 | **Top:** the redundancy ($d\Omega_k$) from the optimal k-multiplet to an H neuron, as found by greedy search, is depicted as a function of time for k ranging from 2 to 10; the plotted quantity is the average over all the H target neurons. **Bottom:** the synergy ($-d\Omega_k$) from the optimal k-multiplet to an H neuron, as found by greedy search, is depicted as a function of time for k ranging from 2 to 10; the plotted quantity is the average over all the H target neurons.

time higher order correlations are dominated by redundancy. The peak at 400 ms is consistent with the peak of TE at 300ms, indeed at time t TE measures the information flow from t to t+1.

Turning to consider take the dynamic transfer of information, we apply the methodology described in section 2.1 and in **Figure 4** we depict the redundancy and the synergy, as a function



of time and for k (the cardinality of the multiplet) ranging from 2 to 10; this figure shows that the response to the stimulus is also shaped by higher-order influences, both of the redundant and synergistic types. Also higher-order influences peak at 300ms, and the synergistic influences seem to show a slower decay after the peak.

In **Figure 5**, we plot the distributions of latencies of maximal redundancy and synergy in optimal multiplets of 10 variables, suggesting that the synergistic response occurs, on average, slightly later than the redundant response.

In **Figure 6**, we depict, as a function of the number of driving variables k , the fraction of variables in the best redundant multiplet belonging to the three classes. We observe that redundant circuits are made of H and M neurons, L neurons rarely appearing in the redundant circuits. On the other hand, we see that L neurons can play a relevant role in synergistic circuits as k becomes larger, and are more important than H and M neurons in the construction of synergistic circuits.

In some instances one may be interested in a particular target neuron and to assess the optimal size of the redundant and synergistic multiplets acting on it. For example, in **Figures 7, 8**, we show how to do it, choosing as the target a randomly selected H neuron. While adding a variable to the redundant multiplet with the greedy search, we also evaluate the redundancy that would be obtained using a randomized version of that variable, and we accept that variable if the probability to get an higher value of the redundancy, after randomization, is less

than 0.05 after Bonferroni correction. For the target neuron under consideration we find that the multiplet with 7 driving variables can be considered statistically significant, as the null hypothesis can be rejected for $k \leq 8$. Here we chose one of the most conservative corrections for multiple comparisons, but the partial dependence across variables could justify more lenient approaches.

In **Figure 8**, we do the same for the synergy using the same target neuron. Since the null hypothesis cannot be rejected at k equal to six, we conclude that the synergistic circuit of five driving variables influencing the target is the largest synergistic multiplet that we can assess statistically.

Some of the figures were generated with Gramm (Morel, 2018).

5. DISCUSSION

We have proposed an approach to analyze higher-order dynamical influences in multivariate time series, and to highlight redundant and synergistic groups of variables influencing a given target variable. Our method generalizes to the dynamic case a recently introduced quantity, named O-information, which was proposed to assess the informational character of equal-time correlations in a set of random variables (Rosas et al., 2019). Our conditioned approach has the main advantage of allowing the distinction of information that is actually exchanged from

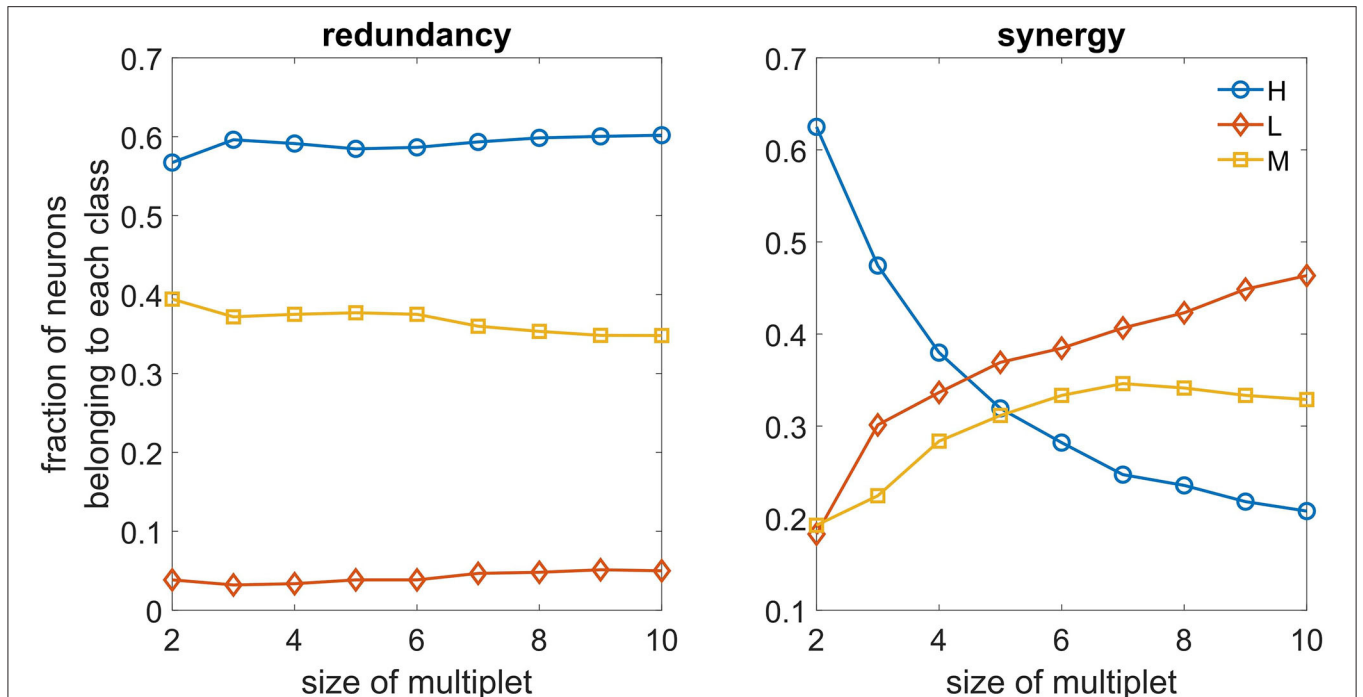


FIGURE 6 | Left: As a function of the size of the optimal multiplet, we depict the typical fraction of H, M, and L neurons constituting the multiplet, both for redundancy (left) and synergy (right).

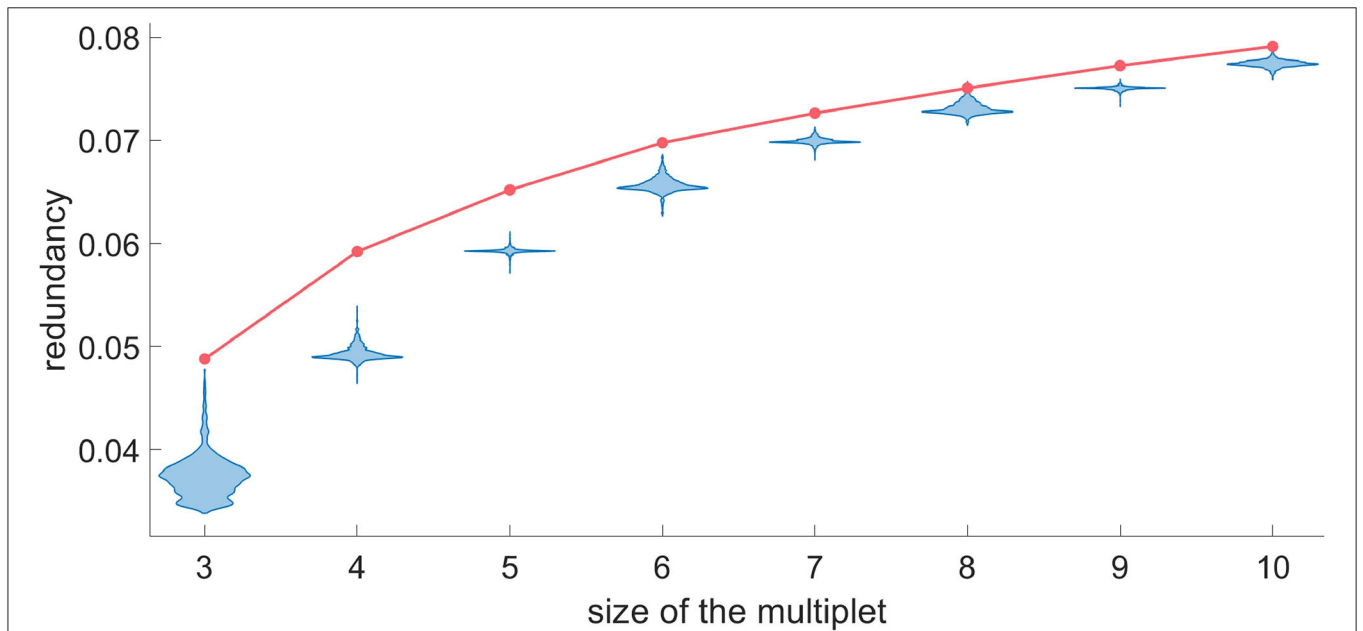
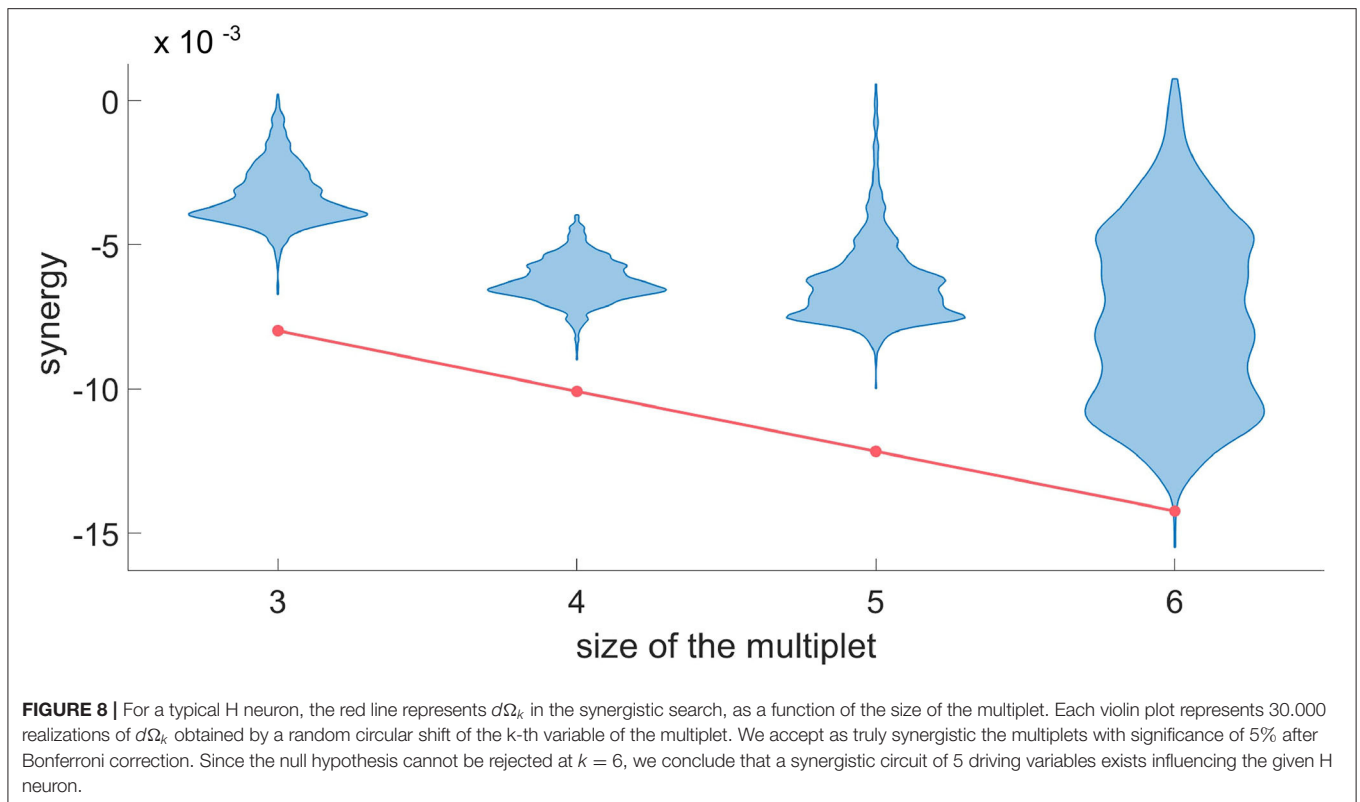


FIGURE 7 | For a representative H neuron, the red line represents the redundancy as a function of the size of the optimal multiplet. Each violin plot represents 30.000 realizations of $d\Omega_k$ obtained by a random circular shift of the k -th variable of the multiplet. We accept as truly redundant the multiplets with significance of 5% after Bonferroni correction. Since the null hypothesis cannot be rejected at $k = 8$, we conclude that a redundant circuit of 7 driving variables exists influencing the given H neuron.

shared information due to common history and input signals. Compared with the expansion in Stramaglia et al. (2012) or PID decompositions in the spirit of Williams and Beer (2010),

the proposed approach is computationally much more feasible. However our approach focuses only on finding multiplets that are synergy-dominated or redundancy-dominated, and the



corresponding values of synergy and redundancy do not come from an exact decomposition of the information flow. For this reason their magnitudes cannot be easily compared for varying k , but in our opinion this is a reasonable price to pay in order to have a fast algorithm that can handle big data sets.

We believe that our approach can have wide applicability in physiology, in particular at the system level where higher-order interactions may play a role in the collective regulation of dynamical rhythms in the human body (Bartsch et al., 2015).

It is worth mentioning that the global transfer entropy of a kinetic Ising model has been shown to have a maximum in the disordered phase (Barnett et al., 2013). Successively it has been shown (Marinazzo et al., 2019) that it is the synergistic component of it that is responsible for this peak, which can be considered as an early warning of a transition toward order. Intuitively, learning processes (storage of new memories by, e.g., Hebbian learning) may be seen as transitions disorder \rightarrow order, and in some sense the response to stimuli described in this paper may be seen as a sequence of transitions disorder \rightarrow order \rightarrow disorder where the control parameter is part of the dynamical process.

The relation between mutual information and synergistic information processing in spiking neurons from organotypic cultures of mouse neocortex was recently addressed in Sherrill et al. (2020), and was found to depend on the timescale and the degree of correlation in neuronal interactions. As an example of application of dO-information, we have considered the response of a neural system to an external stimulus. We have shown that,

in addition to higher order equal time interactions, which show a peak for the redundancy (as probed by the O-information) 400 ms after the go cue, the system displays also significant dynamic transfer of information consisting in synergistic and redundant circuits peaking 300 ms after the go cue. A recent study on computing TE between spiking neurons (Shorten et al., 2020) presented some results on the dependency of the values of TE on the firing rate. Based on these estimations, and given the number target events in the present experiment we can expect that the height of the peak of the TE in **Figure 2** could be slightly overestimated, given the increased firing rate in the same interval. On the other hand the bias is stronger, and toward positive values, with a reduced number of spikes, and the low values before and after the peak are an indication that the TE peak itself is meaningful. The results are further backed up by the surrogate procedure.

Concerning the dynamics of H neurons, from the point of view of pairwise influence, H neurons are the most important drivers, M neurons are also relevant but to a lesser extent, whilst L neurons do not play any role. Going beyond the pairwise description, as far as the redundancy is concerned we find the same relative contributions in terms of the composition of redundant multiplets: the abundance of H neuron is higher than those of M neurons whilst the contribution of L neurons is negligible. On the other hand, considering the synergy the relative importance of the three types of neurons is changed: for large multiplets the abundance of L and M neurons is higher than those of H neurons, thus suggesting that surprisingly also L

neurons may play a role in shaping the dynamics of H neurons by participating in synergistic groups of variables. We have shown that synergy of multiplets of variables can take values up to 0.03 bits. It is worth stressing that dO is not derived from an exact decomposition of transfer entropy and that this value cannot be interpreted as a gain in predictability of the target; however it suggests that the role played by synergistic circuits is small but not negligible when compared with 0.25 bits which is the peak of the global transfer entropy to H neurons, when all the other neurons are simultaneously taken as the driving set. Further investigations are certainly needed to confirm the role of M and L neurons in the higher order description of dynamics of H neurons; our analysis shows that the proposed approach is capable of highlighting these effects while requiring a reasonable computational effort.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: https://github.com/danielemarinazzo/dynamic_O_information.

REFERENCES

- Barnett, L., Barrett, A. B., and Seth, A. K. (2009). Granger causality and transfer entropy are equivalent for gaussian variables. *Phys. Rev. Lett.* 103:238701. doi: 10.1103/PhysRevLett.103.238701
- Barnett, L., Lizier, J. T., Harré, M., Seth, A. K., and Bossomaier, T. (2013). Information flow in a kinetic ising model peaks in the disordered phase. *Phys. Rev. Lett.* 111:177203. doi: 10.1103/PhysRevLett.111.177203
- Barrett, A. B. (2015). Exploration of synergistic and redundant information sharing in static and dynamical Gaussian systems. *Phys. Rev. E* 91:052802. doi: 10.1103/PhysRevE.91.052802
- Bartsch, R. P., Liu, K. K., Bashan, A., and Ivanov, P. C. (2015). Network physiology: how organ systems dynamically interact. *PLoS ONE* 10:e0142143. doi: 10.1371/journal.pone.0142143
- Bashan, A., Bartsch, R. P., Kantelhardt, J. W., Havlin, S., and Ivanov, P. C. (2012). Network physiology reveals relations between network topology and physiological function. *Nat. Commun.* 3, 1–9. doi: 10.1038/ncomms1705
- Bertschinger, N., Rauh, J., Olbrich, E., Jost, J., and Ay, N. (2014). Quantifying unique information. *Entropy* 16, 2161–2183. doi: 10.3390/e16042161
- Bettencourt, L. M. A., Gintautas, V., and Ham, M. I. (2008). Identification of functional information subgraphs in complex networks. *Phys. Rev. Lett.* 100:238701. doi: 10.1103/PhysRevLett.100.238701
- Bossomaier, T., Barnett, L., Harré, M., and Lizier, J. T. (2016). *An Introduction to Transfer Entropy*. Cham: Springer International Publishing. doi: 10.1007/978-3-319-43222-9
- Crutchfield, J. P. (1994). The calculi of emergence: computation, dynamics and induction. *Phys. D Nonlinear Phenomena* 75, 11–54. doi: 10.1016/0167-2789(94)90273-9
- Daniels, B. C., Ellison, C. J., Krakauer, D. C., and Flack, J. C. (2016). Quantifying collectivity. *Curr. Opin. Neurobiol.* 37, 106–113. doi: 10.1016/j.conb.2016.01.012
- Daniels, B. C., Flack, J. C., and Krakauer, D. C. (2017). Dual coding theory explains biphasic collective computation in neural decision-making. *Front. Neurosci.* 11:313. doi: 10.3389/fnins.2017.00313
- Faes, L., and Marinazzo, D. S. (2017). Multiscale information decomposition: Exact computation for multivariate gaussian processes. *Entropy* 19:408. doi: 10.3390/e19080408
- Gatica, M., Cofré, R., Mediano, P. A., Rosas, F. E., Orío, P., Diez, I., et al. (2020). High-order interdependencies in the aging brain. *bioRxiv. [Preprint]*. doi: 10.1101/2020.03.17.995886

ETHICS STATEMENT

The animal study was reviewed and approved by Stanford University Animal Care and Use Committee (IACUC number 9720).

AUTHOR CONTRIBUTIONS

SS and DM contributed to conceptualization and the development of the methodology. SS, TS, and DM contributed to the development of the software and the data analysis. SS, BD, and DM contributed to the design of the experiment. All authors contributed to manuscript revision, read, and approved the submitted version.

FUNDING

SS, Ministero dell'Istruzione, dell'Università e della Ricerca, Award ID: PRIN 2017/WZFTZP.

- Ince, R. A., Giordano, B. L., Kayser, C., Rousset, G. A., Gross, J., and Schyns, P. G. (2017). A statistical framework for neuroimaging data analysis based on mutual information estimated via a gaussian copula. *Hum. Brain Mapp.* 38, 1541–1573. doi: 10.1002/hbm.23471
- James, R. G., Barnett, N., and Crutchfield, J. P. (2016). Information flows? A critique of transfer entropies. *Phys. Rev. Lett.* 116:238701. doi: 10.1103/PhysRevLett.116.238701
- Katz, Y., Tunström, K., Ioannou, C. C., Huepe, C., and Couzin, I. D. (2011). Inferring the structure and dynamics of interactions in schooling fish. *Proc. Natl. Acad. Sci. U.S.A.* 108, 18720–18725. doi: 10.1073/pnas.1107583108
- Kiani, R., Cueva, C. J., Reppas, J. B., and Newsome, W. T. (2014). Dynamics of neural population responses in prefrontal cortex indicate changes of mind on single trials. *Curr. Biol.* 24, 1542–1547. doi: 10.1016/j.cub.2014.05.049
- Kiani, R., Cueva, C. J., Reppas, J. B., Peixoto, D., Ryu, S. I., and Newsome, W. T. (2015). Natural grouping of neural responses reveals spatially segregated clusters in prearcuate cortex. *Neuron* 85, 1359–1373. doi: 10.1016/j.neuron.2015.02.014
- Kiani, R., and Shadlen, M. N. (2009). Representation of confidence associated with a decision by neurons in the parietal cortex. *Science* 324, 759–764. doi: 10.1126/science.1169405
- Lin, A., Liu, K. K., Bartsch, R. P., and Ivanov, P. C. (2020). Dynamic network interactions among distinct brain rhythms as a hallmark of physiologic state and function. *Commun. Biol.* 3, 1–11. doi: 10.1038/s42003-020-0998-x
- Lizier, J., Bertschinger, N., Jost, J., and Wibral, M. (2018). Information decomposition of target effects from multi-source interactions: Perspectives on previous, current and future work. *Entropy* 20, 307–316. doi: 10.3390/e20040307
- Marinazzo, D., Angelini, L., Pellicoro, M., and Stramaglia, S. (2019). Synergy as a warning sign of transitions: the case of the two-dimensional ising model. *Phys. Rev. E* 99:040101. doi: 10.1103/PhysRevE.99.040101
- McGill, W. (1954). Multivariate information transmission. *Trans. IRE Professional Group Inform. Theory* 4, 93–111. doi: 10.1109/TIT.1954.1057469
- Mediano, P. A. M., Rosas, F., Carhart-Harris, R. L., Seth, A. K., and Barrett, A. B. (2019). Beyond integrated information: A taxonomy of information dynamics phenomena. *bioRxiv. [Preprint]*.
- Morel, P. (2018). Gramm: grammar of graphics plotting in matlab. *J. Open Source Softw.* 3:568. doi: 10.21105/joss.00568
- Oihorhenuan, I. E., Mechler, F., Purpura, K. P., Schmid, A. M., Hu, Q., and Victor, J. D. (2010). Sparse coding and high-order correlations in fine-scale cortical networks. *Nature* 466, 617–621. doi: 10.1038/nature09178

- Rosas, F. E., Mediano, P. A. M., Gastpar, M., and Jensen, H. J. (2019). Quantifying high-order interdependencies via multivariate extensions of the mutual information. *Phys. Rev. E* 100:032305. doi: 10.1103/PhysRevE.100.032305
- Schreiber, T. (2000). Measuring information transfer. *Phys. Rev. Lett.* 85:461. doi: 10.1103/PhysRevLett.85.461
- Shadlen, M. N., and Newsome, W. T. (2001). Neural basis of a perceptual decision in the parietal cortex (area lip) of the rhesus monkey. *J. Neurophysiol.* 86, 1916–1936. doi: 10.1152/jn.2001.86.4.1916
- Sherrill, S. P., Timme, N. M., Beggs, J. M., and Newman, E. L. (2020). Correlated activity favors synergistic processing in local cortical networks *in vitro* at synaptically-relevant timescales. *Netw. Neurosci.* 4, 678–697. doi: 10.1162/netn_a_00141
- Shorten, D., Spinney, R., and Lizier, J. (2020). Estimating transfer entropy in continuous time between neural spike trains or other event-based data. *bioRxiv*. [Preprint]. doi: 10.1101/2020.06.16.154377
- Stramaglia, S., Wu, G.-R., Pellicoro, M., and Marinazzo, D. (2012). Expanding the transfer entropy to identify information circuits in complex systems. *Phys. Rev. E* 86:066211. doi: 10.1103/PhysRevE.86.066211
- Wilcoxon, R. (2016). *Introduction to Robust Estimation and Hypothesis Testing, 4th Edn.* Cambridge, MA: Academic Press. doi: 10.1016/B978-0-12-804733-0.00010-X
- Williams, P. L., and Beer, R. D. (2010). Nonnegative decomposition of multivariate information. *bioRxiv*. [Preprint].
- Yu, S., Yang, H., Nakahara, H., Santos, G. S., Nikolić, D., and Plenz, D. (2011). Higher-order interactions characterized in cortical activity. *J. Neurosci.* 31, 17514–17526. doi: 10.1523/JNEUROSCI.3127-11.2011

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Stramaglia, Scagliarini, Daniels and Marinazzo. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.