BioSystems 185 (2019) 104028

Contents lists available at ScienceDirect

BioSystems

journal homepage: www.elsevier.com/locate/biosystems

Using intersection information to map stimulus information transfer within neural networks

Giuseppe Pica[a,1,*], Mohammadreza Soltanipour[a,b,1], Stefano Panzeri[a,*]

[a] Neural Computation Laboratory, Center for Neuroscience and Cognitive Systems @UniTn, Istituto Italiano di Tecnologia, 38068 Rovereto, TN, Italy
[b] Center for Mind/Brain Sciences, University of Trento, 38068 Rovereto, Italy

ARTICLE INFO

Keywords:
Information transmission
Neural coding

ABSTRACT

Analytical tools that estimate the directed information flow between simultaneously recorded neural populations, such as directed information or Granger causality, typically focus on measuring how much information is exchanged between such populations. However, understanding how sensory information is processed through the brain and how it is used to generate behaviors requires estimating specifically the amount of stimulus information that is transmitted. Here we use the concept of intersection information to make progress on how to perform this measure. We develop the concept of transmitted intersection information, which measures how much of the stimulus information present in one population at a certain time is transmitted to a second population at a later time. We show that this measure of stimulus-specific information transfer has several appealing properties, such as being non-negative, and being bounded by the amount of stimulus information present in each of the two populations and by the total amount of information transmitted between the two populations. Applying this measure to simulated neurons or pools of neurons connected by feed-forward synapses, we show that it can discern cases when the information transmitted from one population to another is about specific stimulus features encoded by the sending population from cases in which the information transmitted is not about the stimuli. We also show that this measure has a good statistical sensitivity from trial numbers that can be collected in real data. Our results highlight the promise of using the concept of intersection information to map stimulus-specific information transfer across neural populations.

1. Introduction

The development of tools to monitor simultaneously the activity of populations of neurons has opened up the possibility to understand how information about important external events, such as sensory stimuli, is transmitted across neural populations. Addressing this question is important for many reasons. It is crucial to understand more about the biophysical mechanisms that regulate the transmission and dynamic routing of information across the nervous system. Moreover, it is important to understand how information about the external world travels through the nervous system and generates important behaviors such as the conscious perception of these external stimuli (Van Vugt et al., 2018) or appropriate decisions taken in response to the presence of these external stimuli (Runyan et al., 2017).

Analytical tools derived from the Wiener-Granger principles of causality (Wiener, 1956; Granger, 1969; Bressler and Seth, 2011) and from Information Theory (Massey, 1990; Schreiber, 2000) have been proposed to measure how much information a neural population transmits to another one. These tools for the analysis of simultaneous recordings from multiple neurons or from multiple neural circuits have led to important insights on how information propagates through the nervous system, for example along feed-forward, feedback and lateral connections in cortex (Besserve et al., 2015; Bosman et al., 2012; Van Kerkoerle et al., 2014). However, it is still not known, from the analytical point of view, how to measure not only if some information is transmitted from a population to the next, but also whether the exchanged information is about specific sets or subsets of sensory stimuli, or whether it is about other things, such as internal states. Although some initial attempts have been proposed to address this issue (Ince et al., 2015), much theoretical and conceptual work is still needed in order to develop the tools to map, from neural recordings, stimulus-specific information processing in the brain.

In recent work, we have introduced the new concept of intersection information (Panzeri et al., 2017; Pica et al., 2017b), defined as the

* Corresponding authors.
E-mail addresses: giuseppe.pica@iit.it (G. Pica), m.soltanipour@studenti.unitn.it (M. Soltanipour), stefano.panzeri@iit.it (S. Panzeri).
[1] Co-first author.

https://doi.org/10.1016/j.biosystems.2019.104028
Received 28 February 2019; Received in revised form 5 September 2019; Accepted 6 September 2019
Available online 21 September 2019

information about sensory stimuli in neural activity that is used to generate a behavioral output. This tool is useful to investigate if sensory information in a neural code travels through the readout mechanisms in the brain and contributes to generate an output. In this article, we explore how to extend this recent analytical concept to potentially measure stimulus-specific transmission of information between two neural populations, a sender and a receiver.

2. Material and methods

2.1. Definition and derivation of transmitted intersection information

To measure the information about the stimulus that is transmitted between two neural populations, we generalize the definition of intersection information (I_{II}) that we introduced in Pica et al. (2017b) to study perceptual discrimination. In this Subsection we describe the rationale and the mathematics of the I_{II} measure, with specific application to the problem of neural information transmission.

In a typical stimulus information transmission study, suppose the subject was presented with a discrete stimulus S (for example, the identity of a certain face presented as a visual stimulus). Suppose further that during such presentations we recorded simultaneously activity from two neural populations. Suppose that we quantify the activity of the two populations with two (possibly multidimensional) variables $R1$ and $R2$, which for simplicity we assume to be discrete. For example, the features describing the variables $R1$ and $R2$ could be the total number of spikes of each population in a given trial, discretized in a given number of bins R. We assume that the joint trivariate probability distribution $p(S, R1, R2)$ has been empirically estimated by sampling these variables simultaneously over repeated experimental trials. Shannon's mutual information can be used to estimate the pairwise statistical associations within the set of variables $\{S, R1, R2\}$. $I(S:R1)$ and $I(S:R2)$ measure the sensory information carried about the experimental stimuli by the activity of the first and second neural population, respectively, whereas $I(R1:R2)$ measures the relationship between the activity of the two populations, and can be thus taken as a measure of the total information they exchange. To measure how much information about the stimulus S is routed through the neural feature $R1$ and transmitted to the neural feature $R2$, we propose a novel information-theoretic measure, $I_{II}(S;R1;R2)$. The quantification of this specific information flow involving three stochastic variables requires the use of more sophisticated information-theoretic tools than the pairwise Shannon mutual information. To address this problem, we build on the Partial Information Decomposition (PID) framework (Williams and Beer, 2010; Bertschinger et al., 2014; Harder et al., 2013; Griffith and Koch, 2014), which relies on the decomposition of the mutual information between two source variables and one target variable into four non-negative information components. These components quantify shared (redundant), unique, and complementary (synergistic) modes of information sharing among the three variables, and correspond to finer information quantities than mutual information. To study how stimulus information flows across $R1$ and $R2$, it is convenient to consider the following PID decomposition:

$$\begin{aligned} I(R2{:}\,(S, R1)) = {} & \mathrm{SI}(R2{:}\,\{S; R1\}) + \mathrm{CI}(R2{:}\,\{S; R1\}) \\ & + \mathrm{UI}(R2{:}\,\{S\backslash R1\}) + \mathrm{UI}(R2{:}\,\{R1\backslash S\}), \end{aligned} \tag{1}$$

where

- the shared part $SI(R2:\{S;R1\})$ is the information about $R2$ that we can extract from any of S and $R1$, that is the redundant information about $R2$ shared between S and $R1$;
- the first unique component $UI(R2:\{S\setminus R1\})$ is the unique information about $R2$ that we can only extract from S but not from $R1$. It thus includes stimulus information relevant to $R2$ that is not represented in $R1$;
- the second unique $UI(R2:\{R1\setminus S\})$ is the unique information about $R2$ that we can only extract from $R1$ but not from S;
- the complementary component $CI(R2:\{S;R1\})$ is the information about $R2$ that can only be gathered if both S and $R1$ are simultaneously observed with $R2$, but that is not available when only one between S and $R1$ is simultaneously observed with $R2$. More precisely, it is the part of $I(R2:(S,R1))$ which does not overlap with $I(S;R2)$ nor with $I(R1:R2)$.

The decomposition in Eq. (1) is unequivocally specified after we define the shared information $SI(R2:\{S;R1\})$. Several ways have been defined in the literature to define the shared information (see Pica et al. (2017b)). Here we adopt the following definition (Bertschinger et al., 2014):

$$\mathrm{SI}(R2{:}\,\{S;\,R1\}) \equiv \max_{q\in\Delta_p} \mathrm{CoI}_q(S;\,R1;\,R2), \tag{2}$$

where Δ_p is the space of all probability distributions $q(S, R1, R2)$ such that $q(S, R2) = p(S, R2)$ and $q(R1, R2) = p(R1, R2)$, and $CoI_q(S;R1;R2) \equiv I_q(S:R1) - I_q(S:R1|R2)$ is the co-information evaluated with the probability distribution $q(S, R1, R2)$.

Starting from the above decomposition, we reason that a notion of transmitted information about the stimulus could be the part of the redundant information that S and $R1$ share about $R2$ that is also part of the stimulus information in the first population, $I(S;R1)$. This kind of information is even finer than the information PID components described in the above equations. In previous works (Pica et al., 2017a,b) we showed that comparing information components of different ways to partition the trivariarte system S, $R1$, $R2$ leads to the separation into finer information quantities. Reasoning along these lines, and with a straightforward extensions of (Pica et al., 2017b), we can finally define the transmitted intersection information I_{II} as:

$$I_{\mathrm{II}}(S;\,R1;\,R2) = \min\{\mathrm{SI}(R2{:}\,\{S;\,R1\}),\,\mathrm{SI}(S{:}\,\{R1;\,R2\})\}. \tag{3}$$

It is easy to show, with straightforward extension of Pica et al. (2017b,a), that, with this definition, $I_{II}(S;R1;R2)$ quantifies the part of the redundant information $SI(R2:\{S;R1\})$ that is also a part of $I(S:R1)$. Other desirable properties of $I_{II}(S;R1;R2)$ that follow directly from the definition in Eq. (3) have been described in Pica et al. (2017b). There are a few properties of this measure that are particularly important in what follows and that are summarized here. First, $I_{II}(S;R1;R2) \leq I(S:R1)$, where $I(S:R1)$ is the Shannon information between stimuli and the first neural feature. This is because intersection information is a part of the sensory information carried by the recorded response $R1$, namely, the part which is transmitted to $R2$. Second, $I_{II}(S;R1;R2) \leq I(R1:R2)$, where $I(R1:R2)$ is the Shannon information between the two neural features. This is intuitive because intersection information is a part of such information – namely, the part which is related to the stimulus. Third, $I_{II}(S;R1;R2) \leq I(S:R2)$, where $I(S:R2)$ is the Shannon information between stimuli and $R2$. This is because intersection information is part of the total stimulus information carried by the second neural feature – namely, the part which can be extracted from $R1$.

The numerical computation of $I_{II}(S;R1;R2)$ relies on the numerical solution, through the Franz-Wolfe procedure Frank and Wolfe (1956), of the convex optimization defined in Eq. (2). The convergence of the optimization is theoretically guaranteed because it is convex. However, the numerical implementation of the algorithm converges to a final value that falls close to the true optimal value within some finite (small) tolerance. This tolerance derives from a combination of the tolerance of the Frank-Wolfe optimization (Frank and Wolfe, 1956) and the tolerance of the approximate linear optimizations that are performed at each iteration step. This finite numerical precision of our algorithm to compute intersection information can give, in some cases when the intersection information value is small, Some very small negative values of intersection information (see Fig. 6 for an example). This

problem can in principle be solved by decreasing the tolerance parameters at the cost of a slower computation. Here we set the tolerance parameter to have light computational times on a server, to mimic what would be done in practical analyses of real data. We provided a Matlab implementation of a solving algorithm that is available online https://doi.org/10.5281/zenodo.850362.

2.2. Methods for neural network simulations

To test the new analytical approaches that we wish to develop here, we simulated simple feed forward neural networks in two different scenarios. In one scenario we know that there is stimulus-specific information transfer, and thus we expect to get positive intersection information in this case. In the second scenario there is transfer of information from the first network to the second, but this information is not about the external stimulus and thus we expect to obtain zero stimulus-specific information transfer in this second case.

Our network consists of two layers, and each layer includes 10 leaky integrate and fire neurons. The firing of the input layer 1 is directly modulated by the external stimulus (via the input term ν_{ext}), and neurons in layer 1 are connected to neurons in layer 2 via feed-forward synapses. The neurons in each layer are described as leaky integrated and fire neurons, the membrane potential v of which evolves according to the following equation:

$$\frac{dv}{dt} = \frac{-v}{\tau} + \frac{\nu_{ext}}{\tau_0} + \sigma\sqrt{\frac{2}{\tau}}\chi. \quad (4)$$

When the potential reaches the threshold (without loss of generality the value of threshold is assumed to be one) an action potential occurs, the value of the membrane potential will set to be the rest potential (zero), and the postsynaptic neuron's potential is increased with constant value of (set to 0.3 in unit of threshold). There is no refractory period after each spike. In Eq. (4), τ is the intrinsic time decay constant of the membrane potential, set to 20 ms. $\frac{\nu_{ext}}{\tau_0}$ indicates the external input which could be varied for different stimuli, and is only present for layer 1 neurons. The parameter τ_0 is set to be 10 ms.

The third term in Eq. (4) models the intrinsic noise of the membrane potential with a Gaussian white noise with intensity equal to one, χ, which is multiplied by a parameter σ that modulates the strength of the noise. The value of the noise changes from simulation to simulation and is always reported in unit of threshold in each figure. We note that, in all simulations presented here the network was mean driven since, due to the strong constant input to the layer one (the minimum input for layer one is 2.5 in unit of threshold), absence of inhibitory neurons, and the strong synaptic weight the membrane potential regularly reaches the threshold even in the absence of noise.

We divided layer 1 into two pools of five neurons each (see Fig. 1). Each neuron in layer 2 receives two random excitatory synapses from layer 1, one from each pool. All synapses have the same fixed strength.

We operated the network in two different information coding scenarios (see Fig. 1). In both scenarios, the network can respond differently to two different external stimuli, $s1$ and $s2$.

2.2.1. Scenario 1: transmission of stimulus-related information between the two areas

In the first scenario, all neurons in the first layer respond to the external stimulus in the same way. All $R1$ neurons respond with a high firing rate to stimulus $s1$: $\nu_{ext} = \mu + \Delta S$. All $R1$ neurons respond with a low firing rate to stimulus $s2$: $\nu_{ext} = \mu - \Delta S$. In the above, μ is the mean response across neurons and stimuli (kept to be 4 for all simulations in unit of threshold), and ΔS is a parameter regulating the strength of the stimulus signal (that is, the firing rate differences across stimuli). The higher the ratio $\Delta S/\sigma$, the more information about the stimuli is encoded in the neurons of the first layer.

2.2.2. Scenario 2: transmission between the two areas of information not related to the stimulus

In the second scenario, neurons in the first layer respond to the external stimulus in different ways. The first 5 $R1$ neurons respond with a high firing rate to stimulus $s1$ ($\nu_{ext} = \mu + \Delta S$), whereas the second 5 $R1$ neurons respond with a low firing rate to stimulus $s1$ ($\nu_{ext} = \mu - \Delta S$). The neurons in $R1$ have the opposite response pattern in response to stimulus $s2$: The first 5 $R1$ neurons respond with a low firing rate to stimulus $s2$ ($\nu_{ext} = \mu - \Delta S$), whereas the second 5 $R1$ neurons respond with a high firing rate to stimulus $s2$ ($\nu_{ext} = \mu + \Delta S$). Here it is still the case that the higher the ratio $\Delta S/\sigma$, the more information about the stimuli is encoded in the neurons of the first layer. However, this stimulus information is not transmitted to layer 2 because each neuron of $R2$ receives inputs from one neuron that responds to the stimulus with high firing and another neuron that responds to the stimulus with low firing.

3. Results

3.1. Definition of intersection information for measuring stimulus-specific information transfer within networks

Although the intersection information approach was originally conceived to study how much the sensory information in neural activity r is read out to inform behavioral choice in a perceptual discrimination task (Panzeri et al., 2017; Pica et al., 2017b), the intersection information framework could in principle be extended to study how stimulus information flows across neural populations. In particular, we consider the neural activity of a first brain region $R1$ (whose activity we record) and a downstream area $R2$ (whose activity we assume we can record at the same time when we record $R1$). In this case, we have shown in Methods how to use the intersection information framework to define and compute the quantification of the information about the stimulus S carried by population $R1$ that is transmitted downstream to area $R2$. We have defined in Methods the stimulus specific information transmitted from a first (sender) population to a second (receiver) population as the intersection information $I_{II}(S;R1;R2)$ between the stimulus information in $R1$ and the activity in $R2$, or in other words the amount of stimulus information in $R1$ that is read out by $R2$.

As described in Methods, $I_{II}(S;R1;R2)$ has properties that are highly desirable for a measure of transmission of information about stimuli between two different neural populations. First, $I_{II}(S;R1;R2) \leq I(S:R1)$, where $I(S:R1)$ is the Shannon information about the stimuli carried by neural activity in the first population. Second, $I_{II}(S;R1;R2) \leq I(R1:R2)$, where $I(R1:R2)$ is the Shannon information between the activity in population 1 and the activity in population 2. Third, $I_{II}(S;R1;R2) \leq I(S:R2)$, where $I(S:R2)$ is the Shannon information about the stimuli carried by neural activity in the second population. These properties reflect the intuitive requirements that a good measure of transmission of stimulus-related information should be bounded both by the stimulus information present in each population, and by the total information (about stimuli or other non-stimulus related factors such as internal state variables) transmitted between the two populations.

In the following, we use computer simulations of neural network activity to investigate whether this measure can be used to disambiguate between stimulus-specific and stimulus-unspecific information transmission. The purpose of these simulations was to provide a primary validation of $I_{II}(S;R1;R2)$, and in particular to check the extent to which the proposed measure satisfies some of the basic requirements that we expect of a stimulus-specific information transfer measure. The first expected property is that it is zero in specific cases in which there is no stimulus specific information transfer, even if these cases are contrived. The second expected property is that this information quantity decreases when increasing the amplitude in the noise and increasing when decreasing the amplitude in the stimulus

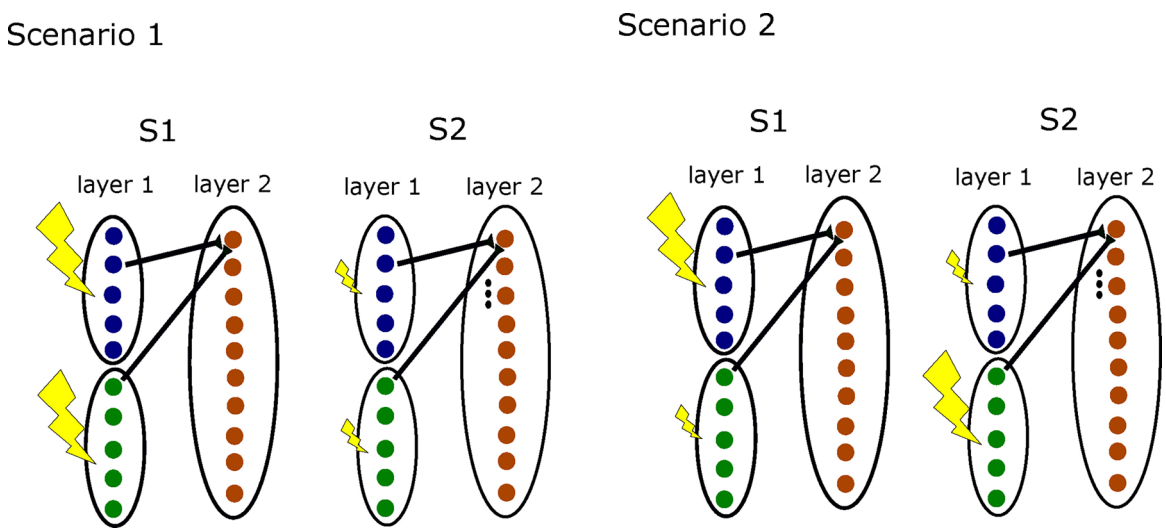

Fig. 1. Schematic of the structure of our simulated network and of the stimulus information flow corresponding to each scenario. Each network has two layers (layer 1, plotted with black and green dots as neurons in this figure; and layer 2, plotted with red dots as neurons in this figure), and each layer has ten neurons. Each neuron in layer 2 receives input from two neurons in layer 1, one from the first pool and one from the second pool of 5 neurons in layer 1. Example inputs from layer 1 to layer 2 are denoted as black lines. The size of the yellow stimulus arrows is proportional to the strength of the response of each pool of layer 1 to each stimulus. Left: in scenario 1 all 10 neurons of $R1$ respond with a more elevated firing rate to stimulus $s1$ and a less elevated firing rate to stimulus $s2$. In this scenario, stimulus information is transmitted from layer 1 to layer 2. Right: in scenario 2, 5 neurons of layer 1 respond with a more elevated firing rate to stimulus $s1$ than to stimulus $s2$, while the other 5 neurons respond with a less elevated firing rate to stimulus $s1$ than to stimulus $s2$. In this scenario, information is transmitted from layer 1 to layer 2, but it is information about the noise and not about the stimulus. Dots in the panel mean that, for simplicity, not all feed-forward connections are shown, but only the ones to the first neuron of layer 2.

signal. For this primary validation, we used a simple feed forward architecture with homogeneous neural properties. The advantage of the proposed architecture is that the effects of these factors are not confounded by other factors that may have a complex and unpredictable effect, such as presence of recurrent loops, feedback loops, heterogeneity, and so on.

3.2. Testing whether intersection information can specifically capture transmission of information about stimuli

Using simulations of a two-layered feed-forward neural network (Fig. 1), we tested the ability of the intersection information $I_{II}(S;R1;R2)$ to disambiguate between cases when there is and there is not transfer of information about the stimuli from the first to the second layer. In these simulations, we ran the network with many trials per stimulus (500 simulated trials for each stimulus). The statistical power of the method, and how this depends on the number of available trials, will be investigated separately in the next Subsection.

We begun by examining the first scenario, in which layer 1 encoded information about the stimuli, and the stimulus information encoded in the activity of layer 1 was passed, though a set of feed-forward synapses, to the neurons of layer 2 (see Methods for a full description).

To compute the information flow, we used the PID decomposition described above. We considered the information $I(S:R1)$ and $I(S:R2)$ about the stimuli carried by a single neuron in the first and second layer, respectively. In other words, we took as $R1$ and $R2$ the firing rate of one neuron in the first and second layer, respectively. We assumed that the responses $R1$ and $R2$ in the two layers were simultaneously recorded in the same trial. In what follows, we will report the average of the information carried by each single neuron in each layer. We also computed the total information $I(R1:R2)$ and the intersection information $I_{II}(S;R1;R2)$ transmitted from a single neuron in layer 1 to a second neuron in layer 2 that was connected to the considered neuron in the first layer. For the information transfer measure, we report the average over all pairs of neurons connected by a synapse (as depicted in Fig. 1). In order to compute numerically the PID quantities, which we derived for a discrete variable case, we discretized the firing rate of each neuron in three equispaced bins.

To compute the time course of the information quantities both in layer 1 and in layer 2, we took a window of duration 10 ms and we slid it along the time course of the neural responses (we however checked that using window durations as short as 4 ms did not change qualitatively the results). To compute information transfer, we considered the activity of one neuron in layer 1 at time t (plotted on the x axis of Fig. 2) and the activity of one neuron in layer 2 with a 5 ms delay, which corresponds to the simulated conduction delay between layer 1 and layer 2. (As a further validation of the soundness of methods, we verified that the information transmission quantities $I(R1:R2)$ and $I_{II}(S;R1;R2)$ had the maximum value when using a 5 ms delay, corresponding to the network conduction delay, to compute the responses $R1$ and $R2$ in the two layers, as expected by simple intuition – data not shown). We report in Fig. 2a,b the time course of the stimulus information of layer 1 and layer 2, respectively, for four representative network parameter values. We found that both layer 1 and layer 2 carried positive stimulus information. Given that layer 2 does not receive a stimulus modulation directly (as shown by the network structure depicted in Fig. 1), the stimulus information in layer 2 must be transmitted from layer 1. To probe the ability of our measure to reveal and quantify directly this information transfer, we computed, for the four representative parameter values chosen, the values of both the total transmitted information $I(R1:R2)$ (Fig. 2c) between the activity of layer 1 and layer 2, and of the stimulus-related intersection $I_{II}(S;R1;R2)$ (Fig. 2d) transmitted between $R1$ and $R2$. We found (Fig. 2d) that $I_{II}(S;R1;R2)$ was positive. Thus our measure correctly identified that there was stimulus information transfer between layer 1 and layer 2, in all cases.

The mathematical properties of $I_{II}(S;R1;R2)$, namely that it is bounded both by stimulus information in layer 1 and in layer 2 and by the total transmitted information (see Methods), make it possible to investigate what fraction of stimulus information in layer 1 is transmitted to layer 2, and also what fraction of the information transmitted between layer 1 and layer 2 is actually about the stimulus. From these comparisons, we found that when noise was lower, the ratio between the transmitted intersection information $I_{II}(S;R1;R2)$ and the stimulus

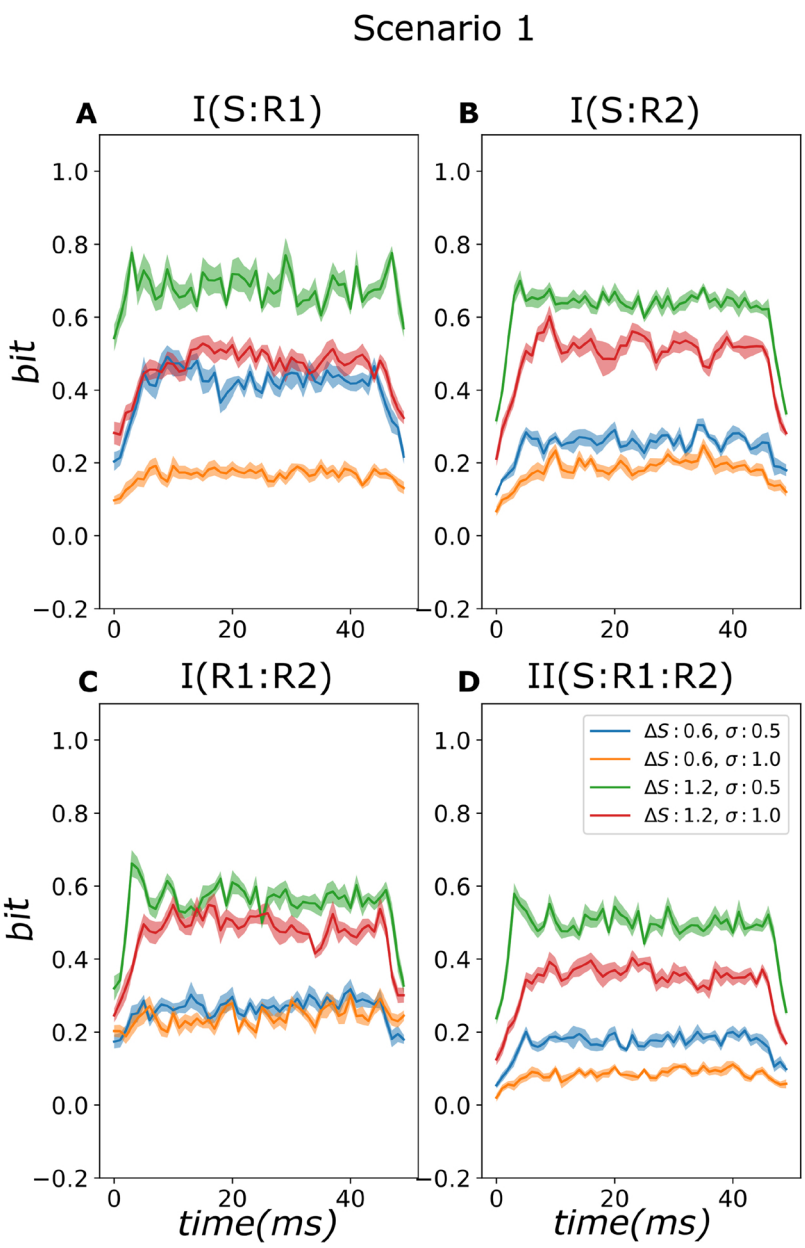

Fig. 2. Information quantities computed for the first scenario using four different combinations of values of the input difference ΔS and noise strength σ parameters. Parameters values are reported in the Legend. Results are plotted as a function of post-stimulus time and reported as mean *pm* sem over 10 simulations with 500 trials per stimulus each. (A) $I(S;R1)$; (B) $I(S;R1)$; (C) $I(R1;R2)$; (D) $II(S;R1;R2)$.

information $I(S:R1)$ in layer 1 was very high, close to 1. This means that, under such conditions, all stimulus information in layer 1 was transmitted. When noise was higher, the ratio $I_{II}(S;R1;R2)/I(S:R1)$ was lower than 1, especially in the case of lower stimulus modulation (orange line in Fig. 2), showing that less of the stimulus information was transmitted, due to the higher noise in layer 1. These results suggest that when noise is increased, higher proportion of the total information $I(R1:R2)$ transmitted from the first to the second layer is about the noise. This is indeed what we found, as the ratio $I_{II}(S;R1;R2)/I(R1:R2)$ was close to 1 (meaning that all transmitted information was about the stimulus) for lower noise and was considerably less than 1 (meaning that a considerable part of the transmitted information was about noise fluctuations and not about the stimulus signal) for higher noise level.

To further appreciate how the values of intersection information, as well as those of the other information quantities, depend upon the parameters of the network, we ran systematic simulations by varying parametrically the stimulus signal ΔS and noise (of the first layer) σ parameters. Results of these simulations are reported in Fig. 3. In this case, given that the results in Fig. 2 show that the information quantities are relatively constant in time, we present the values of the information quantities averaged over the entire 0-50 ms post-stimulus time. We found that the stimulus information in both layers ($I(S;R1)$ in Fig. 3a and $I(S;R2)$ in Fig. 3b), as well as the intersection information $I_{II}(S;R1;R2)$, increased when increasing ΔS and/or decreasing σ (see Fig. 3d). This was expected since the increase of signal and the decrease of noise should increase the available stimulus information and how it is transmitted. The comparison between the intersection information and the values of $I(S;R1)$ and of $I(R1;R2)$ allowed us to explore the fraction of the stimulus information in layer 1 and the fraction of the total information transmitted across layers that are turned into transmitted intersection information. We found the trend that we highlighted when considering Fig. 2 held also with the more systematic exploration of the network's parameter space considered in Fig. 3.

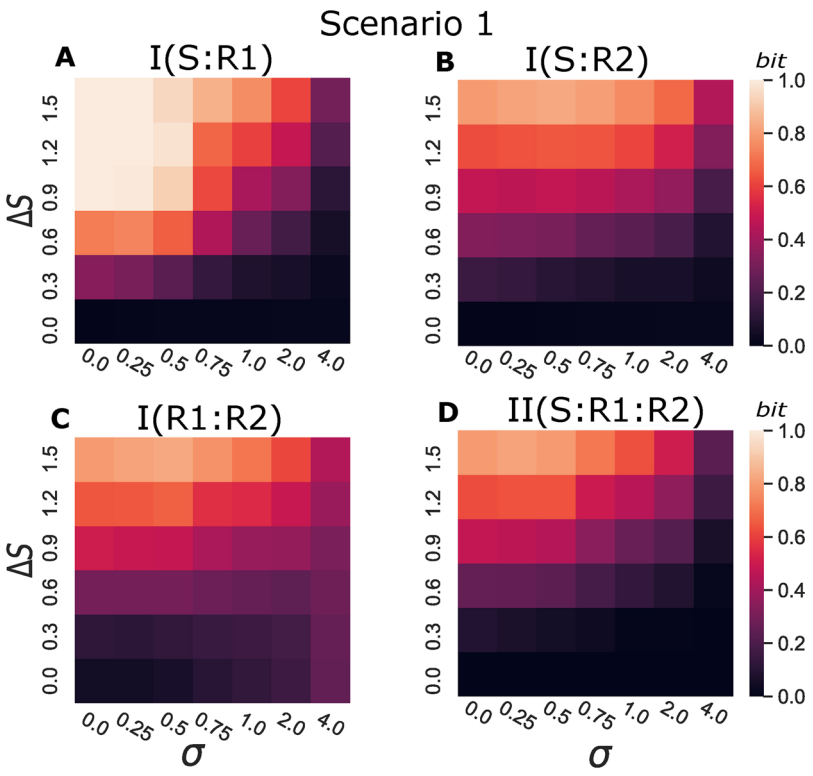

Fig. 3. Information quantities computed for the first scenario using different four different combinations of values of the signal strength ΔS and noise strength σ parameters. Parameters values are reported in the Legend. Results are computed for neural responses averaged over 50 ms post-stimulus time (considering 5 ms propagation delay for $R2$) and reported as mean pm sem over 10 simulations with 500 trials per stimulus each. (A) $I(S;R1)$; (B) $I(S;R1)$; (C) $I(R1;R2)$; (D) $II(S;R1;R2)$.

It is interesting to note that in Figs. 2 and 3, for some network parameter values, we occasionally found cases in which the value of $I(S:R2)$ was higher than that of $I(S:R1)$. Given that layer 2 receives stimulus information only through layer 1, this result seems at first to contradict the data processing inequality. However, this inequality would necessarily be true only if $I(S:R1)$ included all neurons in Layer 1 (thus all sources of stimulus information to Layer 2). Given that in this figure we consider as feature $R1$ the firing rate of a single neuron in layer it does not need to be true in all cases that $I(S:R2)$ is limited by $I(S:R1)$. In intuitive terms, this is because neural response feature $R1$ does not capture all sources of stimulus information to neural response $R2$.

To mimic the case in which in which a mass signal that collects activity from many neurons but that cannot resolve them individually, such as the measure of a multi-unit activity in each area, we extended the computations presented in the previous two figures to include the analysis of the pooled firing rate response in each layer. (The responses $R1$ and $R2$ were still one-dimensional arrays, but contained pooled multi-unit activity rather than activity of just one neuron as in the previous figures). We varied parametrically the number of neurons per layer that were pooled together from one (as in the analyses presented in the previous two figures) to five. Results are reported in Fig. 4. We found that all quantities increased with the number of neurons, but that the pattern of results across measures and simulation parameters was very similar to the one observed in previous Figures, thereby confirming also in this case that the intersection measures meets the requirements that we expected.

To further validate these measures, we performed a new set of simulations for scenario 1 when we increased both the size of the network and the number of connections per neuron. Results of information values at the center of the plateau (25 ms post-stimulus) are reported in Fig. 5. We kept the number of pooled neurons per layer fixed to five.

We found that, in general, having more neurons and a higher percentage of connections increased all types of information, probably because of a larger averaging of noise in the larger or more connected network. Importantly, also in this case all the pattern of results across measures and simulation parameters was very similar to the one observed in previous figures, thereby further confirming that the intersection measure meets the expected requirements.

It is fundamental to check if the transmitted intersection information measure is able to disambiguate cases in which transmitted information is about the stimulus from cases when transmitted is not about the stimulus, we simulated a second scenario in which neurons in layer 1 still encoded positive stimulus information as in the first scenario, and still transmitted information to layer 2, but did not transmit information about the stimuli (see Methods and Fig. 1 right panel). (We note that we verified that in this case there was no transmission of information about the stimuli neither by signal nor by noise, by checking that both the mean and the standard deviation of firing rates of neurons of Layer two were significantly modulated by the stimuli, t test $P > 0.3$). This scenario is admittedly contrived, but it is important to explore whether transmitted intersection information would be null in this case. We found (see Fig. 6) that for all parameter values that we investigated stimulus information in the first layer was still significant, and transmitted information from layer 1 to layer 2 was still high, but the transmitted intersection information and the information in layer 2 were zero. (Note that these results held for a much wider range of noise values, including small noises, than those reported in Fig. 6). These results are important because they suggest that the transmitted intersection information formalism is suitable to distinguish between cases when information transmission is about the stimulus from other cases. Or in other words, these results confirm that the transmitted intersection information really focuses on transmission of stimulus information, thus corroborating our theoretical considerations.

3.3. Testing the statistical sensitivity of the transmitted intersection information

In the previous section we tested the ability of transmitted intersection information to disambiguate between stimulus-specific and stimulus-unspecific information transmission scenarios. We thus concentrated on simulations with relatively large trials numbers (500 trials

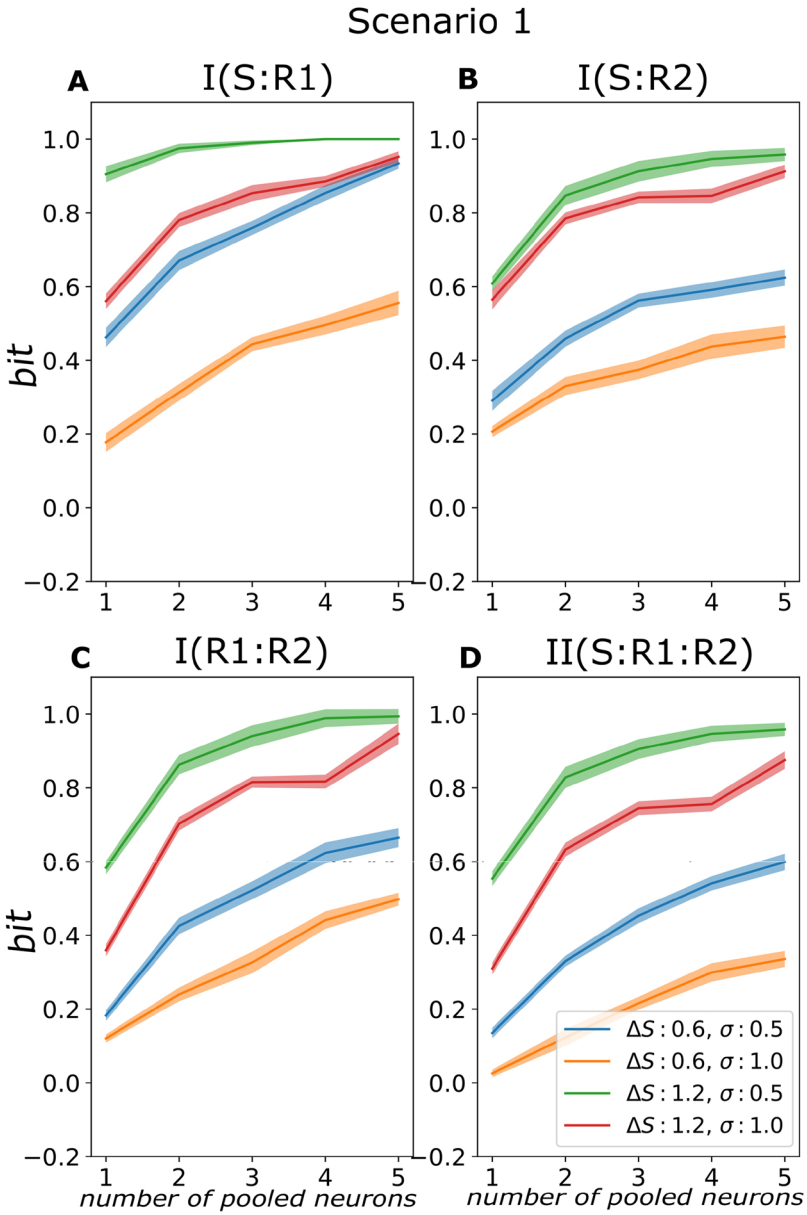

Fig. 4. Information quantities as a function of the number of pooled neurons, computed at the post-stimulus time $t = 25$ ms. Results are reported as mean *pm* sem over 10 simulations with 500 trials per stimulus each. (A) $I(S;R1)$; (B) $I(S;R1)$; (C) $I(R1;R2)$; (D) $II(S;R1;R2)$.

per stimulus, which is in the upper limit of what may be available experimentally, which ranges from few tens to few hundreds trials per stimulus). However, in some application to real data, the number of available trials per stimulus may be scarce. An important practical question is what is the statistical power of our measure $I_{II}(S;R1;R2)$, and namely how the ability of the measure to detect even small values of transmitted stimulus information scales with the number of available trials. Here we focus only on the statistical power of intersection information, because the statistical power of mutual information was already extensively studied in Ref. Ince et al. (2012). In this section we focus on measures of transmitted intersection information based on one neuron per layer, like those reported in Figs. 2–4.

We first constructed a non-parametric permutation test of the null hypothesis that there is no intersection information. This null hypothesis was constructed by randomly permuting (without replacement) the values of $R1$ across trials while leaving untouched the values of S and $R2$. This way, the information between S and $R2$ is left untouched, but it cannot be mediated by $R1$ any more. Thus, intersection information must be null in such permuted datasets. We then used the null hypothesis distribution of transmitted intersection information values to detect whether a value of intersection information computed from the network with a fixed number of trials was significantly larger than zero with a threshold p value $p < 0.05$. We ran several different simulations with different noise strengths, in order to obtain a range of different (albeit small) values of intersection information. We concentrated on small values because, for obvious reasons, we expected that the sensitivity of the measure would be stronger when intersection information is larger. The sensitivity (defined as the fraction of measures in which the presence of non-zero intersection information was correctly detected in simulations with positive ground-truth intersection information, i.e. in scenario 1) is reported as function of the number of trials per stimulus in Fig. 7. We found that the sensitivity was very high (0.7 or more) even for very small values of intersection information (in the range 0.01 to 0.03 bit) for as little as 100 trials per stimulus, which is well in the range of experimental values even in awake animals (Runyan et al., 2017) when extensive data collection is more challenging. These results suggest that the measure has a sufficient sensitivity to be applied to real data.

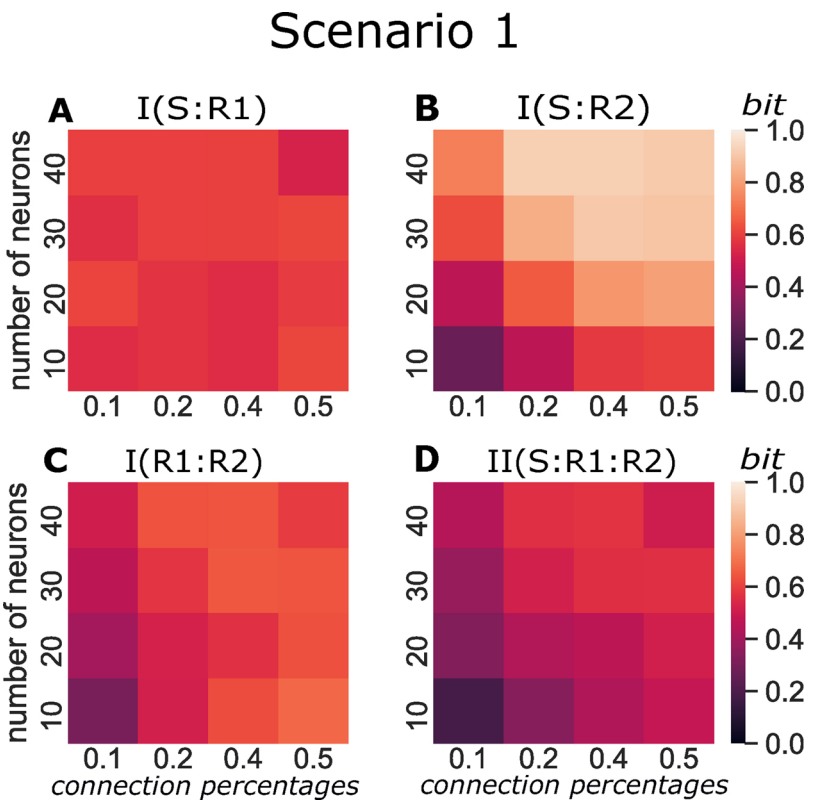

Fig. 5. Information quantities as a function of number of neurons per layer and percent of connection between layer one and layer two for Scenario one. ΔS and σ are kept fixed with the values of 0.6 and 1.0 measures are calculated at the post-stimulus time $t = 25$ ms. Results are reported as mean *pm* sem over 10 simulations with 500 trials per stimulus each. (A) $I(S;R1)$; (B) $I(S;R1)$; (C) $I(R1;R2)$; (D) $II(S;R1;R2)$.

3.4. Investigating the limited sampling bias of the transmitted intersection information

Previous work has shown that discretized estimators of Shannon information quantities suffer from an upward limited sampling bias, which can be corrected with simple methods such as quadratic or linear extrapolations of their dependence on the sample size (Panzeri and Treves, 1996; Steven et al., 1998; Panzeri et al., 2007). The results of the studies of the properties of the discrete estimators of Shannon information is that (especially for larger numbers of bins, which is the case where the bias is most concerning) the Shannon information quantity (among the 3 Shannon information quantities that we compute, $I(S:R1)$, $I(S:R2)$, $I(R1:R2)$) is $I(R1:R2)$. For this latter quantity, the analytical approximation to the bias in the large sample size regime scales approximately as $(R-1)2/2N(ln2)$, where N is the total number of trials, and R is the number of bins for both $R1$ and $R2$ (Panzeri and Treves, 1996). Thus, the bias grows quadratically with the number of bins used to discretize the neural responses in each layer. (In case we consider only one neuron or one pooled neural population per layer, as we do here, R would be equal to the number of bins used to discretized this single neural feature. In case we consider n simultaneously recorded neuron per layer, R would be equal to the power n of number of bins used to discretized each neural feature). This previous work showed that this bias can be corrected and subtracted out using linear or quadratic extrapolations of the scaling of information with the number of trials very effectively as long as N is at least 4–5 times larger than R Panzeri and Treves (1996), Steven et al. (1998), Panzeri et al. (2007). This means that, if we use a range of number of bins R per layer from 2 to 10, we would expect that $I(R1:R2)$ is accurately computed if 20 to 400 trials are available in total across all stimuli.

However, the scaling properties with sample size of $I_{II}(S;R1;R2)$, which being derived from an optimization procedure within the PID, may be different from those a directly computed Shannon information, and are not yet known. Here we investigated and reported (Fig. 8) the scaling of transmitted intersection information $I_{II}(S;R1;R2)$ as function of the number of trials per stimulus and the number of bins used to compute $R1$ and $R2$, after applying a bias correction based on a linear extrapolation of the value of the estimator with sample size Steven et al. (1998). (We focused on transmitted intersection information computed from one neuron per layer, like those reported in Figs. 2–4.) We used data generated from network simulations obtained under scenario 1. As in the simulations performed in Panzeri et al. (2007), the extent to which the information estimator is data robust can be studied by considering the smallest number of trials by which the information reaches its asymptotic value that is reached for larger trial numbers. We found that the calculation of $I_{II}(S;R1;R2)$ reached stably its asymptotic value, for all values of numbers of bins in each layer that we varied from 2 to 10) when using approximately 20–50 trials per stimulus. (This range is well within the range of total trials used in real experiments, which goes from a hundred or so trials for experiments with behaving subjects (Runyan et al., 2017) to several hundred trials for experiments under anaesthesia Arabzadeh et al. (2004)). We concluded that transmitted intersection information was well computed when Shannon information between Layer 1 and Layer 2 was well computed, and that transmitted intersection information $I_{II}(S;R1;R2)$ is no more cursed by the dimensionality problem than the total Shannon information exchanged between the two layers.

4. Discussion

Recent advances in neuroscience technology allow the simultaneous recording from multiple neurons in multiple regions during behavioral tasks. This allows an unprecedented opportunity to dissect and reverse engineer the information transmission pathways that are required for accurate performance of behavioral tasks (Guo et al., 2014; Peng et al., 2015; Daniel et al., 2013).

In this paper we introduced and explored the properties of a new analytical measure, that we termed transmitted intersection information, that could be used as a mathematical tool to analyze simultaneous recordings from multiple brain regions. The primary validation of this

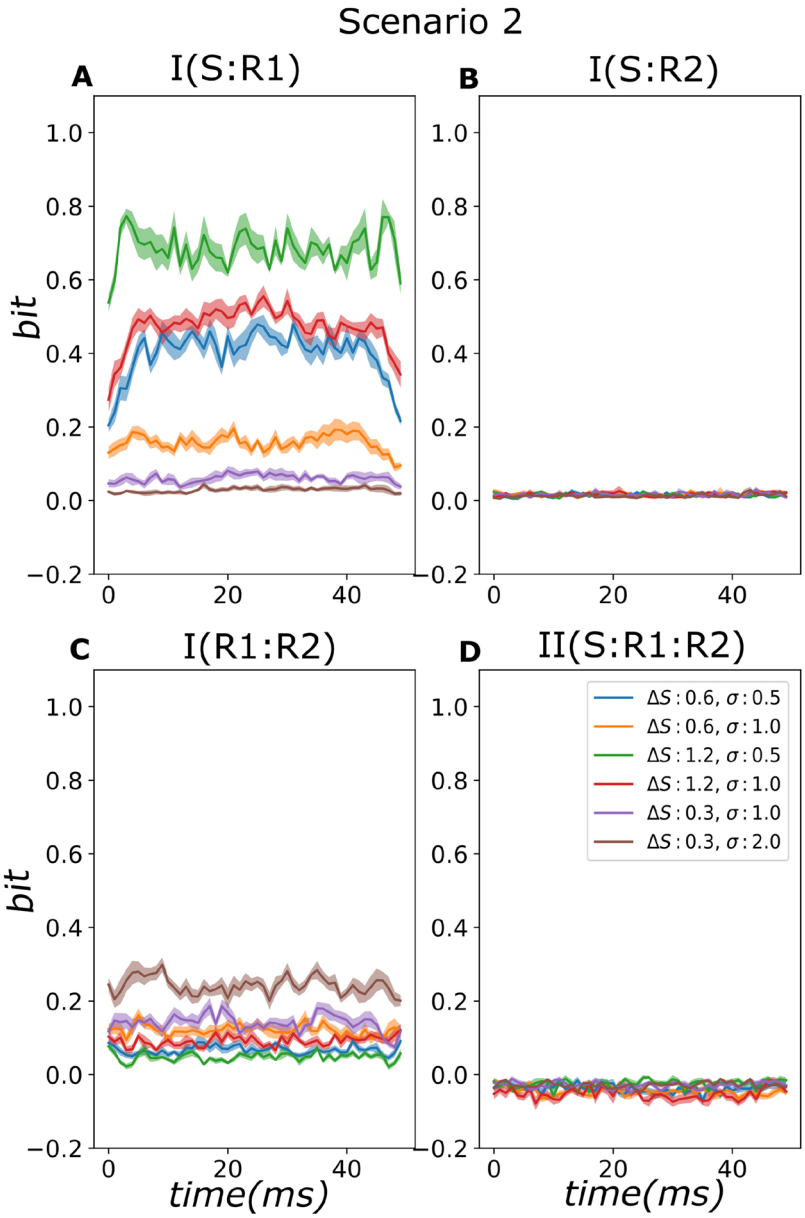

Fig. 6. Information quantities computed for the second scenario using six different combinations of values of the input difference ΔS and noise strength σ parameters including the values used in Fig. 2. Parameter values are reported in the Legend. Results are plotted as a function of post-stimulus time and reported as mean *pm* sem over 10 simulations with 500 trials per stimulus each. (A) $I(S;R1)$; (B) $I(S;R1)$; (C) $I(R1;R2)$; (D) $II(S;R1;R2)$.

method on neural network simulations was successful in indicating that the method can pick genuine transfer of information about a set of stimuli, and that it satisfies some basic characteristics expected of this measure, including that it should increase with increasing stimulus signal and it should decrease with increasing stimulus noise in the information transmission. This tool could be used to identify the neural response features that transmit specific information to downstream regions, leading to hypotheses about the mechanisms of information flow in neural circuits. It could also be used to help identify the neural pathways that carry specific information about external correlates and dissect their function.

The method can be applied to situations in which neural activity is recorded simultaneously from multiple brain regions. Here we validated the method in simulated cases in which only one neuron or one mass population signals (multi-unit activity) was measured from each brain regions considered. Although such simulations are important to establish easily important primary validations of this method, future simulations involving the analysis of many single neurons for each area would be valuable to understand how these ideas can be extended to explore the role of redundancies and synergies within areas due to interactions between neurons. For practical reasons, performing this analysis would require further work to understand how to couple most effectively the formalism to data compression techniques, which we leave to further studies.

When simultaneous recordings from multiple brain regions are performed during perceptual discrimination tasks, this formalism could be used to better understand the role in perceptual discrimination performance of communication pathways among circuits. The ability of the transmitted intersection information measure to discriminate between stimulus-specific and stimulus-unspecific transmission of information could be in principle used to determine whether a projection pathway carries information that is instructive (Otchy et al., 2015) for task performance (contributes essential information for the task performance) or if the circuit is permissive for task performance (modulates the behavior but does not provide essential information). The former case would require that the considered pathway carries

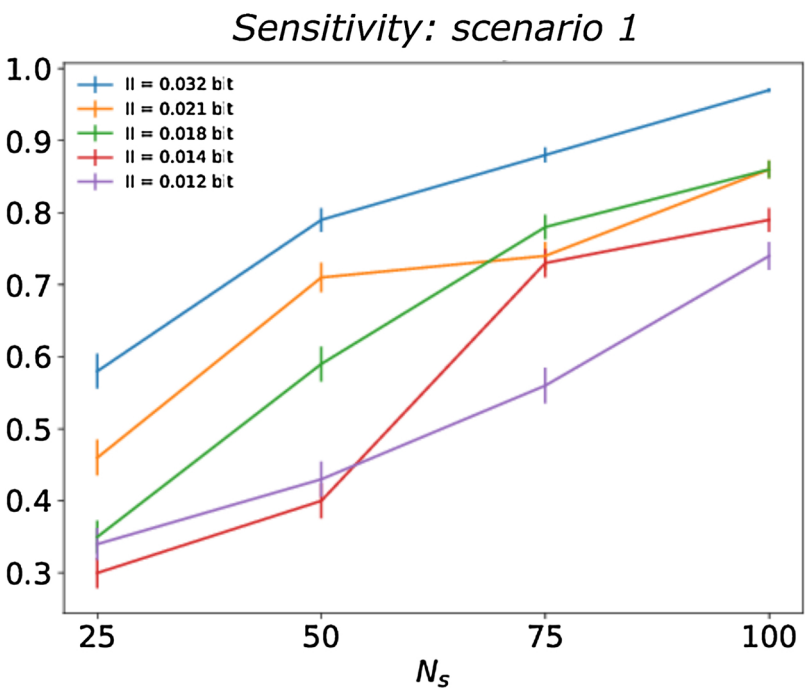

Fig. 7. Sensitivity of the Transmitted Intersection Information $I_{II}(S;R1;R2)$ measure as a function of the number of simulated trials per stimulus N_S. For five different values of the noise strength parameter and for each number of trials we ran 100 independent realizations of the simulation with scenario 1. For each simulation, we built a bootstrap distribution of 40 $II_{sh}(S;R1;R2)$ values by randomly shuffling across trials the instances of $R1$. Sensitivity was computed as the fraction of simulated $I_{II}(S;R1;R2)$ values that were larger than the 95th percentile of the corresponding bootstrap distribution. Error bars were constructed on the assumption that the significance of $I_{II}(S;R1;R2)$ is Bernoulli distributed.

information about the stimulus features that are needed for performing the task. The latter case may show information transmitted between the nodes of the pathway, but this information does not need to be about the stimuli to be discriminated in the task.

Given that it is becoming possible to couple, during behavior, the use of methods for the large scale recording of neural activity with methods for precise perturbation of neural activity (for example combining imaging with optogenetics (Packer et al., 2014; Emiliani et al., 2015)), one feature of the intersection information approach that may become of particular importance is the fact that it naturally lends itself to being generalized for the use with perturbation experiments. The transmitted intersection information formalism would allow to design experiments in which patterns of neural activity that carry specific stimulus signals could be imposed in one layer, and then a readout of information about the stimulus features whose signature was imposed with optogenetics could be extracted from another neural population, and the causal information flow between the two could be quantified using causal extensions of transmitted intersection information.

Another measure has been recently proposed, from one of us and other colleagues, to identify stimulus specific information transfer from neural recordings (Ince et al., 2015). In this previous work, the stimulus specific information was defined as redundancy between stimulus representations in the two layers at different times. This measure was shown to be statistically powerful and useful to determine patterns of information flow in the human brain (Ince et al., 2015). However, in that previous work redundancy was defined as the difference between the sum of the information carried by two variables separately and the information they carry jointly, and this definition actually conflates redundant and synergistic effects. This problem implies that the earlier measure can become negative, and thus more difficult to interpret, when the relationship between the two layers is dominated by synergistic effects. The use of the PID, that we introduced in this work, allows us to alleviate this problem, by isolating the redundant information and discarding the synergistic effects. As a result, our measure $I_{II}(S, R1, R2)$ is always positive and thus can be easily interpreted, and is measured in units of bits which have a in intuitive meaning. Another advantage of using the PID is that, as discussed in Methods, the transmitted intersection information is bounded by the stimulus information encoded in each layer and the total information transmitted across layers. These bounds, as we discussed in Results, have the potential to help the experimenter to determine how efficient is the transmission of stimulus information across populations.

The Wiener – Granger principle that is used for the statistical analysis of causal information flow requires that we assess whether present activity in the second node can be predicted from past activity at the first node above and beyond how present activity in the second node can be predicted from the past of the second node itself. With transmitted intersection information, so far we have assessed whether stimulus information in the second node at the present time is shared with the stimulus information present in the first node at the earlier time. However, to fully implement the spirit of the Wiener – Granger principle, we would need to make sure that the stimulus information shared between the present of the second node and the past of the first node is novel, or unique, with respect to the information present in the past of the second node. This problem is extremely difficult to deal with Partial Information Decomposition approaches, as it requires considering an additional stochastic variable (the past of the second node) in the information decompositions. Indeed, it would need identifying not only, as we do now, information about the activity in the first layer that is shared both by the stimulus and the second layer, but also information about the stimulus in the second layer that is unique with respect to the stimulus information that was present in the second layer at earlier times. However, the number of partial information terms in a PID diagram grows very rapidly with the number of stochastic variables that we analyze (Williams and Beer, 2010), and the PID of a four-variable system already consists of 18 partial information terms. It would be extremely challenging to estimate reliably all such terms based on the trial numbers that are commonly used in neuroscience experiments. We thus believe that implementing in full these principles would require coupling our theory with good and data-robust parametric (Sheikhattar et al., 2018; Francis et al., 2018) or non parametric models (Safaai et al., 2018) of the relationships between stimuli and neural activity. We plan to pursue this avenue in future work.

However, for the time being we note that the present formalism can be applied already in full satisfaction of the Wiener – Granger principle if we apply it to ask the question of whether the first wave of stimulus information present in the second node comes from the stimulus information present at an earlier time in the first node. This is because if

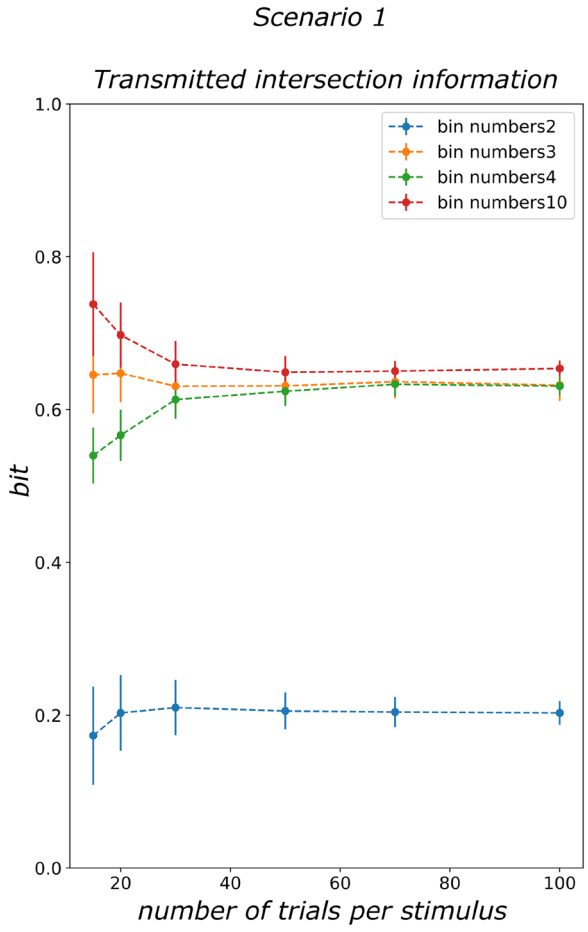

Fig. 8. Values of the estimator of Transmitted Intersection Information $I_{II}(S;R1;R2)$ as a function of number of trials per stimulus N_S, calculated for different values (from 2 to 10) of the number of bins R used to compute $R1$ and $R2$. We used network simulations of scenario 1 with ΔS and noise strength parameter equal to 1.2 and 0.25 respectively. For each value of number of trials per stimulus preorted on the x axis, we ran 10 independent realizations of the simulation. Results plot mean ± sem of $I_{II}(S;R1;R2)$ over these 10 realizations.

we concentrate on the first wave of information in the second layer, we know that no other stimulus information was present in the second layer at earlier times. Therefore the results presented here are useful both to explore the power of the concept of intersection information and of PIDs to reverse engineer information transmission in real neural systems, and to already analyze, in its present form, specific parts of datasets.

Conflict of interest

The authors declare no conflict of interest.

References

Arabzadeh, E., Panzeri, S., Diamond, M.E., 2004. Whisker vibration information carried by rat barrel cortex neurons. J. Neurosci. 24 (26), 6011–6020.
Bertschinger, N., Rauh, J., Olbrich, E., Jost, J., Ay, N., 2014. Quantifying unique information. Entropy 16 (4), 2161–2183.
Besserve, M., Lowe, S.C., Logothetis, N.K., Schölkopf, B., Panzeri, S., 2015. Shifts of gamma phase across primary visual cortical sites reflect dynamic stimulus-modulated information transfer. PLOS Biol.
Bosman, C.A., Schoffelen, J.M., Brunet, N., Oostenveld, R., Bastos, A.M., Womelsdorf, T., Rubehn, B., Stieglitz, T., De Weerd, P., Fries, P., 2012. Attentional stimulus selection through selective synchronization between monkey visual areas. Neuron 75 (5), 875–888.
Bressler, S.L., Seth, A.K., 2011. Wiener-granger causality: a well established methodology. Neuroimage 58 (2), 323–329.
Daniel, H., O'connor, S., Andrew Hires, Zengcai, V., Nuo Li, G., Yu, J., Sun, Q.-Q., Huber, D., Svoboda, K., 2013. Neural coding during active somatosensation revealed using illusory touch. Nat. Neurosci. 16 (7), 958.
Emiliani, V., Cohen, A.E., Deisseroth, K., Häusser, M., 2015. All-optical interrogation of neural circuits. J. Neurosci. 35 (41), 13917–13926.
Francis, N.A., Winkowski, D.E., Sheikhattar, A., Armengol, K., Babadi, B., Kanold, P.O., 2018. Small networks encode decision-making in primary auditory cortex. Neuron 97 (4), 885–897.
Frank, M., Wolfe, P., 1956. An algorithm for quadratic programming. Nav. Res. Logist. Q. 3 (1–2), 95–110.
Granger, C.W.J., 1969. Investigating causal relations by econometric models and cross-spectral methods. Econom.: J. Econom. Soc. 424–438.
Griffith, V., Koch, C., 2014. Quantifying synergistic mutual information. Guided Self-Organization: Inception. Springer Berlin Heidelberg, pp. 159–190.
Guo, Z.V., Li, N., Huber, D., Ophir, E., Gutnisky, D., Ting, J.T., Feng, G., Svoboda, K., 2014. Flow of cortical activity underlying a tactile decision in mice. Neuron 81 (1), 179–194.
Harder, M., Salge, C., Polani, D., 2013. Bivariate measure of redundant information. Phys. Rev. E 87 (1), 012130.
Ince, R.A.A., Mazzoni, A., Bartels, A., Logothetis, N.K., Panzeri, S., 2012. A novel test to determine the significance of neural selectivity to single and multiple potentially correlated stimulus features. J. Neurosci. Methods 210, 49–65.
Ince, R.A.A., Van Rijsbergen, N.J., Thut, G., Rousselet, G.A., Gross, J., Panzeri, S., Schyns, P.G., 2015. Tracing the flow of perceptual features in an algorithmic brain network. Sci. Rep. 5, 17681.
Massey, J., 1990. Causality, feedback and directed information. In: Proc. Int. Symp. Inf. Theory Applic. (ISITA-90). Citeseer. pp. 303–305.
Otchy, T.M., Wolff, S.B.E., Rhee, J.Y., Pehlevan, C., Kawai, R., Kempf, A., Gobes, S.M.H., Olveczky, B.P., 2015. Acute off-target effects of neural circuit manipulations. Nature 528, 358–363.
Packer, A.M., Russell, L.E., Dalgleish, H.W.P., Häusser, M., 2014. Simultaneous all-optical manipulation and recording of neural circuit activity with cellular resolution in vivo. Nat. Methods 12 (2), 140.
Panzeri, S., Treves, A., 1996. Analytical estimates of limited sampling biases in different information measures. Netw. Comput. Neural Syst. 7 (1), 87–107.
Panzeri, S., Senatore, R., Montemurro, M.A., Petersen, R.S., 2007. Correcting for the sampling bias problem in spike train information measures. J. Neurophysiol. 98 (3), 1064–1072.
Panzeri, S., Harvey, C.D., Piasini, E., Latham, P.E., Fellin, T., 2017. Cracking the neural code for sensory perception by combining statistics, intervention, and behavior. Neuron 93 (3), 491–507.
Peng, Y., Gillis-Smith, S., Jin, H., Tränkner, D., Ryba, N.J.P., Zuker, C.S., 2015. Sweet and bitter taste in the brain of awake behaving animals. Nature 527 (7579), 512.
Pica, G., Piasini, E., Chicharro, D., Panzeri, S., 2017a. Invariant components of synergy, redundancy, and unique information among three variables. Entropy 19 (9), 451.
Pica, G., Piasini, E., Safaai, H., Runyan, C., Harvey, C., Diamond, M., Kayser, C., Fellin, T., Panzeri, S., 2017b. Quantifying how much sensory information in a neural code is relevant for behavior. Adv. Neural Inf. Process. Syst. 3686–3696.
Runyan, C.A., Piasini, E., Panzeri, S., Harvey, C.D., 2017. Distinct timescales of population coding across cortex. Nature 548 (7665), 92.
Safaai, H., Onken, A., Harvey, C.D., Panzeri, S., 2018. Information estimation using nonparametric copulas. Phys. Rev. E 98 (5), 053302.
Schreiber, T., 2000. Measuring information transfer. Phys. Rev. Lett. 85 (2), 461.
Sheikhattar, A., Miran, S., Liu, J., Fritz, J.B., Shamma, S.A., Kanold, P.O., Babadi, B., 2018. Extracting neuronal functional network dynamics via adaptive granger causality analysis. Proc. Natl. Acad. Sci. USA 115 (17), E3869–E3878.
Strong, S.P., Koberle, R., de Ruyter van Steveninck, R.R., Bialek, W., 1998. Entropy and information in neural spike trains. Phys. Rev. Lett. 80 (1), 197.
Van Kerkoerle, T., Self, M.W., Dagnino, B., Gariel-Mathis, M.-A., Poort, J., Van Der Togt, C., Roelfsema, P.R., 2014. Alpha and gamma oscillations characterize feedback and feedforward processing in monkey visual cortex. Proc. Natl. Acad. Sci. USA 111 (40), 14332–14341.
Van Vugt, B., Dagnino, B., Vartak, D., Safaai, H., Panzeri, S., Dehaene, S., Roelfsema, P.R., 2018. The threshold for conscious report: signal loss and response bias in visual and frontal cortex. Science 360 (6388), 537–542.
Wiener, N., 1956. The theory of prediction. Modern Mathematics for Engineers.
Williams, P., Beer, R., 2010. Nonnegative Decomposition of Multivariate Information. arXiv:1004.2515.