Determining what information is transmitted across neural populations

Jan Bím

Supervisor:

Prof. Stefano Panzeri

Neural Computation Lab, Center for Neuroscience and Cognitive Systems, Istituto Italiano di Tecnologia
Center for Mind/Brain Sciences (CIMeC), University of Trento

The dissertation is submitted for the degree of
Doctor of Philosophy

December 2017

ABSTRACT

Quantifying the amount of information communicated between neural population is crucial to understand brain dynamics. To address this question, many tools for the analysis of time series of neural activity, such as Granger causality, Transfer Entropy, Directed Information have been proposed. However, none of these popular model-free measures can reveal what information has been exchanged. Yet, understanding what information is exchanged is key to be able to infer, from brain recordings, the nature and the mechanisms of brain computation. To provide the mathematical tools needed to address this issue, we developed a new measure, exploiting benefits of novel Partial Information Decomposition framework, that determines how much information about each specific stimulus or task feature has been transferred between two neuronal populations. We tested this methodology on simulated neural data and showed that it captures the specific information being transmitted very well, and it is also highly robust to several of the confounds that have proven to be problematic for previous methods. Moreover, the measure was significantly better in detection of the temporal evolution of the information transfer and the directionality of it than the previous measures. We also applied the measure to an EEG dataset acquired during a face detection task that revealed interesting patterns of interhemispheric phase-specific information transfer. We finally analyzed high gamma activity in an MEG dataset of a visuomotor associations. Our measure allowed for tracing of the stimulus information flow and it confirmed the notion that dorsal fronto-parietal network is crucial for the visuomotor computations transforming visual information into motor plans. Altogether our work suggests that our new measure has potential to uncover previously hidden specific information transfer dynamics in neural communication.

ACKNOWLEDGEMENTS

To Stefano Panzeri, my supervisor: he gave birth to the research question behind the project and the idea of how to solve it. He always pointed in the correct direction on the way to achieve the goal and kept inspiring and motivating me both directly and by his scientific achievements. Last, but not least, I would like to express gratitude to him for creating an environment which was not only very supportive for my research but also allowed me to manage and advance other aspects of my life.

Special thank you to colleagues who cooperated on the project. To Andrea Brovelli for his work on the MEG experiments and countless consolations regarding all aspects of the thesis. To Daniel Chicharro for introducing me into some of the aspects of the underlying mathematics and for always being open to discuss the mathematical parts of the project. To Vito De Feo for helpful discussions and to Eugenio Piasini for his helpful comments to the method and for helping to broaden my horizons in fields that I was a novice in.

I am grateful to my whole family, who were always there for me, during the whole studies, created an environment in which I could fully concentrate on my studies and remained supportive and providing further energy, despite the physical distance between us, whenever I stumbled on the path that ultimately lead to this thesis.

With a special mention to IIT, CIMeC and University of Trento for creating the opportunity to work on the project.

And finally, last but by no means least, also to everyone in the whole institute and around for creating very pleasant, family-like atmosphere that was very supportive.

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION

Cognitive functions arise from the dynamic coordination of neural activity, both across neurons and populations within a local circuit and across large-scale networks (Bressler & Menon, 2010; Panzeri, Macke, Gross, & Kayser, 2015; Varela, Lachaux, Rodriguez, & Martinerie, 2001; von der Malsburg, Phillips, & Singer, 2010). Tracking how neural populations and brain regions interact and exchange information, and what information they exchange, is therefore crucial for modern neuroscience. It is, in particular, crucial for being able to infer from brain activity recordings both the nature of neural computations made by the brain and the putative circuit mechanisms that lead to these computations.

There are many examples that serve to illustrate the importance of establishing how neural populations communicate and what they communicate. An important example is the study of resting state networks in humans and animals from functional magnetic resonance imaging (fMRI) recordings. These studies (Biswal, Zerrin Yetkin, Haughton, & Hyde, 1995; Dosenbach et al., 2007; Fox & Raichle, 2007; Fox et al., 2005; Raichle, 2015) have revealed consistent patterns of brain co-activation when a human subject is at rest, and this has led to important results about default modes of communications across key brain structures and to many important and yet unresolved questions about how information flows across the whole brain both at rest and when performing specific tasks (Buckner, Andrews-Hanna, & Schacter; Greicius, Krasnow, Reiss, & Menon, 2003; Greicius, Supekar, Menon, & Dougherty, 2009; Liska, Galbusera, Schwarz, & Gozzi, 2015; Lu et al., 2012). Another example includes the study of the flow of different kinds of visual information across the ventral and cortical dorsal stream in primates and humans (Aggleton, Keen, Warburton, & Bussey, 1997; Andermann, Kerlin, Roumis, Glickfeld, & Reid, 2011; DiCarlo, Zoccolan, & Rust, 2012; Gallardo et al., 1979; McDaniel, Coleman, & Lindsay, 1982; Tafazoli et al., 2017). Researchers

have found that different features of visual information, such as "what" and "when" information, seem to each travel and be processed along a separate pathway (the "when" information in the dorsal pathway, the "what" information in the ventral pathway). Moreover, researchers have used information about the anatomical connections, the progressive latency of visual responses across cortical areas, and the changes in neural response tuning to objects to infer putative information processing pathways along each stream. For example, in the ventral stream scientists have individuated a series of areas, culminating with inferior temporal cortex in monkeys, with progressively longer response latencies to complex images (such as objects and faces) and progressively increasing invariance of responses to transformations of these objects (Alemi-Neissi, Rosselli, & Zoccolan, 2013; DiCarlo & Cox, 2007; Rosselli, Alemi, Ansuini, & Zoccolan, 2015; Vinken, Vermaercke, & de Beeck, 2014; Zoccolan, Oertelt, DiCarlo, & Cox, 2009). These facts have led to elaborate theories, partly influenced by machine learning and partly influencing new developments in machine learning (Cadieu et al., 2007; DiCarlo et al., 2012; Hassabis, Kumaran, Summerfield, & Botvinick, 2017; Riesenhuber & Poggio, 1999; Serre, Oliva, & Poggio, 2007; Serre, Wolf, Bileschi, Riesenhuber, & Poggio, 2007), of how information is transmitted and is transformed across ventral brain areas. However, the exact circuit mechanisms that lead to this putative information flow across the visual system remain largely undetermined and will need to be addressed with simultaneous recordings from neural populations across several stations of the ventral pathway, coupled with tools that can established exactly what specific aspects of visual information is being passed from one population to the next. Other important open problems in neuroscience that require being able to map information flow in neural circuits include understanding how information flows across different laminae of the cortical circuit. It has been long known that cortical microcircuits exhibit a layer organization that is found across many areas, and it has long been hypothesized that this organization is key to performing canonical cortical computations (Bastos

et al., 2012; Callaway & Marder, 2012; Douglas, Martin, & Whitteridge, 1989). However, to understand if and how these canonical computations are performed, it is necessary to be able to identify and measure the flow of specific information across the canonical microcircuit (Bosman et al., 2012; Klink, Dagnino, Gariel-Mathis, & Roelfsema, 2017; Lamme & Roelfsema, 2000; Van Kerkoerle et al., 2014).

Due to the importance of understanding the flow of information across neurons and populations of neurons, there have been many studies attempting to develop and use mathematical techniques for the analysis of simultaneous neural recordings have been developed and used for this issue. As reviewed e.g. in (Horwitz, 2003), the first technique that was used to study the neuronal interaction was cross correlation. It was used both with electrophysiological recordings from multiple units (AM Aertsen, Gerstein, Habib, & Palm, 1989; AMHJ Aertsen & Preissl, 1991; Gerstein & Perkel, 1969) and with time series obtained from EEG (Adey, Walter, & Hendrix, 1961; Barlow & Brazier, 1954; Gevins et al., 1985). Later, the Granger Causality (Granger, 1969) was adopted to describe putative causal information transfer between time series of brain recordings. For example, it has been used to study information flow between neural activity captured by different EEG sensors (Brovelli et al., 2004; Kamiński, Ding, Truccolo, & Bressler, 2001), or to study information flow between neural activity captured by recordings of fMRI activity over many brain regions (Roebroeck, Formisano, & Goebel, 2005; Sato et al., 2006), or finally to study information flow across laminae of the cerebral cortex (Bosman et al., 2012; Van Kerkoerle et al., 2014; Van Kerkoerle, Self, & Roelfsema, 2017). Concurrently, approaches based on Shannon's Information Theory (Shannon, 1948), have been developed in the dynamical systems and time-series analysis community to measure information flow across time series . Such measures include Directed Information (DI) (Amblard & Michel, 2011; Massey, 1990) and a similar quantity – Transfer Entropy (Schreiber, 2000). It has also been adopted in neuroscience, e.g. on electrophysiological signals (Pereda, Quiroga, & Bhattacharya, 2005) or the domain

of fMRI (Lizier, Heinzle, Horstmann, Haynes, & Prokopenko, 2011). Other model-based approaches to estimate information flow include Dynamic Causal Modelling (DCM) (Friston, Harrison, & Penny, 2003). These methods are based on fitting a specific neural circuit model to the joint time series of neural activity from multiple populations, and then to infer communication flow from the model parameters and the model structure that best fit the data.

A problem with the above-mentioned methods of measuring or inferring information flow from time series of neural activity is that these measures are good to quantify if different neural populations exchange information and how much information they exchange. However, a major problem of these existing methods is that they cannot quantify *what* information has been exchanged. In the example of the ventral visual system detailed above, these methods would for example be able to tell if area A sends information to area B, but they would not be able to tell whether area A is sending for example information about the shape of the face or its color, or whether this information is sent in a transformation-invariant or in a transformation independent way.

To address this void in the literature, which is key to make progress on understanding circuit operations from brain recordings, in this thesis we develop a measure that quantifies explicitly how much information area A sends to area B about specific stimulus features, such as those spelled out in the example above. We call the measure rDFI because it builds on an idea of measure called Directed Feature Information (DFI) (Ince et al., 2015) that was created, based on the Shannon's Information Theory (Cover & Thomas, 2012; Shannon, 1948), with the intent to address the same problem. However, we redefined it as a redundancy (thus rDFI) in scope of the novel Partial Information Decomposition framework of (Williams & Beer, 2010). rDFI exactly measures how many bits of information about a particular stimulus

feature were transferred from an area A to an area B. This thesis is organized as follows. In chapter 2 we briefly review existing methods for the quantification of how much information has been exchanged by neural populations from the analysis of the time series of their joint recordings. In chapter 3 we extensively test the new methods, and compare it to existing methods. We test the methodology both with extensively numerical simulations that express different scenarios for information transmission, and with EEG and MEG recordings. The EEG dataset is a dataset acquired during a face detection task that revealed interesting patterns of interhemispheric phase-specific information transfer. In the MEG dataset of visuomotor associations we analyzed high gamma activity. Our measure allowed for tracing of the stimulus information flow and it confirmed the notion that the dorsal fronto-parietal network is crucial for the visuomotor computations transforming visual information into motor plans. Finally, in chapter 4 we discuss the possible implications of this work for brain research.

CHAPTER 2: A BRIEF REVIEW OF EXISTING METHODS FOR QUANTIFYING INFORMATION FLOW ACROSS NEURAL POPULATIONS

Here we present a succinct review of the two most widely used model-free statistical methods to quantify effective connectivity, highlighting both their strengths and how they need to be improved.

The first, and perhaps the most used measure of putative causality from time series is Granger causality (GC). Granger causality builds on the causality definition of (Wiener, 1956) that was operationalized by (Granger, 1969), and has been widely adopted by the neuroscientific community (Brovelli et al., 2004; Kamiński et al., 2001; Roebroeck et al., 2005). The most common implementation of Granger causality is modelling the two random variables (neuronal signal) using linear autoregressive process models and comparing the coefficients of dependence of one process on another or as shown by (Geweke, 1984) computing the logarithm of the ratio of respective residuals of those processes. Granger causality from time series X to time series Y, in intuitive terms, measures how much our ability to predict the present value of time series Y is increased, with respect to what we can predict only based on the past of Y, when we know the past of time series X. Thus, Granger causality quantifies how much the past of X influences the present value of Y after discounting the autocorrelation of Y. Granger causality has a rather intuitive definition, readily-available means of implementation and it is well established as an analysis method within the community (Bressler & Seth, 2011). However, it suffers from certain limitations. Granger causality by its definition only captures linear interactions in the data and requires relatively low levels of noise in the data (Nalatore, Ding, & Rangarajan, 2007). Moreover, and importantly as far as this Thesis work is concerned, GC measures putative causal interactions or information transfer from X to Y, but it cannot tell if the transmitted information is about a certain stimulus or task related features or not. For

example, it could be that area X transmits to area Y only internal state information that has nothing to do with the task or stimulus variables of interest. Granger causality would not be able to distinguish this possibility from the possibility that instead X is transmitting information about a specific stimulus or task relevant variables to Y, and it would not be able to determine which features' information is transmitted and which is not.

Transfer Entropy (Schreiber, 2000) and Directed Information (Massey, 1990) are rigorous implementations of the Wiener causal principle that make use Shannon's information theory (Shannon, 1948). Assuming that there is no instantaneous transfer between the neurons and using only one value for the past for Directed Information instead of a sum across all past values, the two measures; Transfer Entropy and Directed Information, are equivalent (Amblard & Michel, 2011). Since these are indeed assumption of this work, we will consider only one of these measures in the remaining text.

Focusing on Directed Information, the definition of this quantity is described next. Consider two time series, with a common time domain, X and Y (for the purpose of this thesis they are time series of simultaneously recorded neurophysiological signals but the definition is in principle general). Directed Information from X to Y is then defined as the reduction of uncertainty of the present value Y_t of Y, given the knowledge of its past, by the knowledge of the past of X.

$$(2.1)\quad DI(X \to Y) = H(Y_t \mid Y_{past}) - H(Y_t \mid Y_{past}, X_{past})$$

Where $H(Y_t \mid Y_{past})$ quantifies the uncertainty of Y given its past defined as the Shannon's conditional entropy.

$$(2.2)\quad H(Y \mid X) = -\sum_{x \in X} p(x) \sum_{y \in Y} p(y \mid x) \log_2 (p(y \mid x))$$

For completeness, the unconditional entropy used in the next equation is then

$$(2.3)\quad H(X) = -\sum_{x \in X} p(x) \log_2 (p(x))$$

It can be shown that Directed Information can be reformulated as a mutual information between the past of X and the present of Y conditioned on the past of Y.

$$(2.4)\quad DI(X \to Y) = I(Y_t; X_{past} \mid Y_{past})$$

Where $I(Y_t; X_{past} \mid Y_{past})$ is the Shannon's mutual information, which is defined as

$$(2.5)\quad I(X;Y) = H(X) - H(X \mid Y)$$

and the conditional mutual information is defined followingly:

$$(2.6)\quad I(X;Y \mid Z) = H(X \mid Z) - H(X \mid Y,Z)$$

This definition of Directed Information is perhaps more intuitive for understanding of Directed Information as a measure of information transfer given that entropy can be seen as a measure of information in a time series. Then Directed Information from X to Y is defined as the amount of information shared between the past of X and the present of Y that is novel with respect to the past of Y. However, neither (eq. 2.1) nor (eq. 2.4) are usually sufficient for estimating Directed Information in practice because of the potential complexity of estimations of the probability distribution of the full past of X and Y. Therefore, it is customary to limit the pasts to a low number of values or to choose a single value in the past that occurred with a certain delay of d time points (Schreiber, 2000), resulting into the following definition:

$$(2.7)\quad DI(X \to Y) = I(Y_t; X_{t-d} \mid Y_{t-d}).$$

Unlike Granger causality, Transfer Entropy and Directed Information have the advantage that they rely on the full probabilities of the time series (see above equations) and thus do not make any a priori assumption about the form of the interaction between the two time series, and therefore they are potentially able to capture the effect of arbitrarily complicated types of coupling. However, even these information measures have certain limitations. The limitations come with the fact that it is in practice difficult to sample the full joint probabilities of the time series with all possible lags between them. Thus, these quantities are usually estimated incorporating a number of parameters (like the delay in the above equations and time lags in Granger Causality) that must be set "ad-hoc" by the data analyst and have the potential to significantly influence the outcome of the method. Moreover, these quantities require larger amounts of data than Granger Causality to be used and this restricts its application range, in particular when multivariate extensions are considered (Vicente, Wibral, Lindner, & Pipa, 2011). Finally, like Granger Causality, also Transfer Entropy and Directed Information suffer from the problem that they cannot reveal what information is being transmitted.

In summary, even though Granger Causality, Transfer Entropy and Directed Information can measure the amount of information transmitted from one neuronal population to another, they do not cast any light on what do the neuronal populations communicate to each other. The closest one of the methods to understand the content of the transmitted information, mentioned earlier, is Dynamic Causal Modelling. For the way it fits the circuit models to the data, it can detect stimulus dependent information transfer. However, none of its current implementations can distinguish which stimulus features were exactly communicated. Also, it suffers from assumptions that might be hard to satisfy and therefore limit its wide use across different scenarios.

It is quite intriguing that there was done a significant amount of work on understanding the content and the coding of information in neuronal responses e.g. (Haxby, Connolly, & Guntupalli, 2014; Kriegeskorte, Goebel, & Bandettini, 2006; Panzeri et al., 2015) but literature on understanding the content of communication between neuronal populations is rather scarce.

The first preliminary proposal to address this issue was conceived by Stefano Panzeri and was published by Ince and collaborators (Ince et al., 2015). The authors proposed a conditional measure of DI, called Directed Feature Information (DFI), which quantifies features-specific (e.g., stimulus-type) information transfer between neural signals (Ince et al., 2015). This measure will be defined mathematically in the next section. However, and as we will show in detail by examples in the next chapter, the DFI suffers from several severe conceptual problems. The first is that the DFI can be negative, therefore suggesting that it might not capture the exact notion of information transfer since the meaning of a negative transfer is not clear. DFI can be reformulated as an instance of a measure called Interaction Information (McGill, 1954) or co-information (Bell, 2003), which negative values are considered to emerge from a synergy and positive values from a redundancy being present in the system. It was shown that the standard Information Theory cannot capture the notion of those effects of multivariate interaction between variables. It considers only an interaction between two variables at a time and incorporates others only in conditioning that is aimed to discount their effect on the original pair (Williams & Beer, 2010). After studying some previous concepts trying to generalize the standard Information Theory to multivariate interactions between variables such as Total Correlation (Watanabe, 1960) (used also under the name Multivariate Constraint (Garner, 1962), Multiinformation (Studený & Vejnarová, 1998) and Integration (Tononi, Sporns, & Edelman, 1994)), and approaches based on correlations (Panzeri, Schultz, Treves, & Rolls, 1999; Pola, Thiele, Hoffmann, & Panzeri, 2003), Williams and Beer proposed a method based on a lattice composed of non-negative partial

information terms (Paul L Williams and Randall D Beer 2010). Their method solves the general problem of dissecting a mutual information of a target and its set of predictors into nonnegative terms representing unique, synergistic and redundant information separately (explained in detail later). Since it can be shown that the information transfer can be formulated as a pure redundancy, this should allow for a precise expression of the intended information transfer. The aim of this thesis is to develop a new measure of information transfer that can distinguish between transferred stimuli and that is based on the novel framework of (Williams & Beer, 2010).

CHAPTER 3: DERIVATION OF A NEW MATHEMATICAL MEASURE OF WHAT INFORMATION IS EXCHANGED BY POPULATIONS OF NEURONS

In this chapter, we will describe the derivation of our new measure. We consider two time series, X and Y, representing the activity of two neuronal populations or brain regions, and a random variable S representing the experimental (dependent or independent) variable we wish to study. The feature S may be a stimulus type, a behavioral response or a task condition. Our goal was to develop a model-free measure quantifying information transfer from X to Y that is about S. We reasoned that for such transfer to occur there needs to be certain set of conditions satisfied. First, the sender of an information, X in our case, has to contain that information at a past point in time. Second, also the receiver of the information, Y, has to contain the information at a later point in time than the sender. Third, there needs to be an information transfer between the sender and the receiver present at a point in time between the presence of the information in X and its presence in Y. Finally, the information that is present in both X and Y must be the same information and it cannot be present in Y at any time before the information transfer, otherwise the transfer is not the origin of the information.

Here we summarize and formalize the idea. For a transfer of information about S from X to Y the following predicates must be satisfied (Fig. 1):

1. The past of X (at time t_1) must carry information about S.
2. The present of Y (at time t_2) must carry information about S.
3. Information about S must be transferred from X to Y between t_1 and t_2
4. The past of X and the present of Y must share information about S that was not already present in the past of Y.

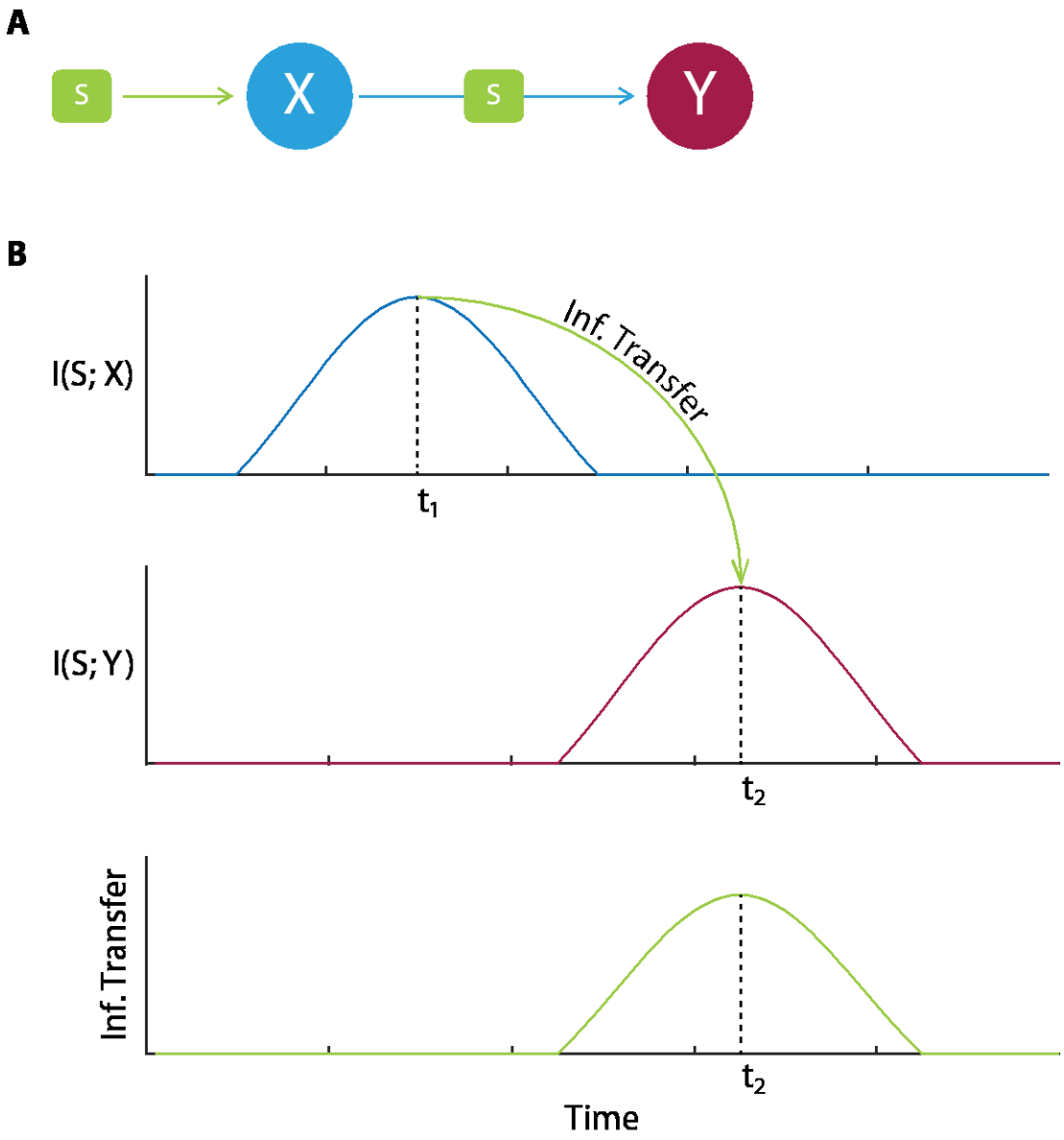

Figure 1 - We reasoned that four key predicates must be satisfied in order for information transfer about a certain stimulus S from X to Y to occur ***A****. The mutual information between S and X or Y (blue and red line in* ***B****) confirms satisfaction of the first two predicates; that in a past time t1 there was information about S in X and that in the present (marked as t2) there is information about S in Y. The green line then determines satisfaction of the third predicate by confirming an information transfer between X and Y that is aligned with the presence of the information about S. There is no visualization of the fourth predicate because the measure that could represent it is the subject of development in this thesis.*

The first predicate is satisfied if there is an increase in the mutual information between S and the past of X, $I(S; X_{t_1})$, as the mutual information guarantees presence of information that X and S share

or in other words that X contains about S and vice versa. The second holds true if there is an increase in $I(S;Y_{t_2})$ because of the same logic as in the first predicate. The third predicate is satisfied if the Directed Information (DI), defined in (eq. 2.7), between the X and Y is non-zero. As a reminder, DI is the information theoretic measure quantifying reduction in uncertainty about the value of Y gained from observation of past values of X, that cannot be obtained by observing the past values of Y itself (Massey, 1990). As shown in (Amblard & Michel, 2011; Ince et al., 2015), DI is equivalent to Transfer Entropy (Schreiber, 2000) in our conditions. The fourth predicate, however, remains challenging. Note, that a candidate measure that satisfies the fourth predicate will most likely be indicative of the first three too. To address the challenge of the fourth predicate, a measure called Directed Feature Information (DFI) was introduced to quantify such information transfer:

$$(3.1)\quad DFI^{S}_{X\to Y} = DI_{X\to Y} - (DI_{X\to Y} \mid S)$$

DFI expresses the difference between the total information transfer between X and Y ($DI_{X\to Y}$) and the information transfer when the stimulus S is known ($DI_{X\to Y} \mid S$). The intuition behind the formulation is that the first term quantifies all information that was transferred between X and Y and the second one quantifies all information that was transferred while S is known and therefore the information is not about S. Hence, after the subtraction, the remaining information is the transfer between X and Y about S because all other transfer information was subtracted. In summary, DFI quantifies the amount of new information transferred from X to Y that is related to the variations in S (Ince et al., 2015).

In fact, DFI can also be expressed as a co-information (Bell, 2003; McGill, 1954), or a redundancy Red_S, between X and Y at given time points with respect to S:

(3.2) $DFI^{S}{}_{X\to Y} = \mathrm{Re}\, d_S(X_{t_1}; Y_{t_2} \mid Y_{t_1'}) = I(S; X_{t_1} \mid Y_{t_1'}) + I(S; Y_{t_2} \mid Y_{t_1'}) - I(S; X_{t_1}, Y_{t_2} \mid Y_{t_1'})$

Where $t_1 < t_2$ and $t'_1 < t_2$. Note, that past of Y can be taken from a different time point than the past of X, and that it is a 3-variables problem with fixed S. DFI quantifies the difference between information provided by the past of X and the present of Y when observed separately and together. Due to synergistic effects, a phenomenon that allows for the total information provided by two variables to be higher when observed together than when observed separately, DFI can become negative. Despite the fact that DFI was successfully applied to detect stimulus dependent transfer (Ince et al., 2015), its property of potentially being negative hinders its interpretation as the amount of information transfer.

For better understanding of causes and implications of what was stated directly above, consider an information interaction of a system of two variables X and Y in the scope of the Shannon's Information theory (Fig. 2). The total amount of information in the system is quantified by the joint entropy $H(X,Y)$

.

(3.3) $H(X,Y) = H(X \mid Y) + H(Y)$

Where $H(X \mid Y)$ and $H(X)$ are a conditional entropy and an entropy as defined in (eq. 2.2) and (2.3).

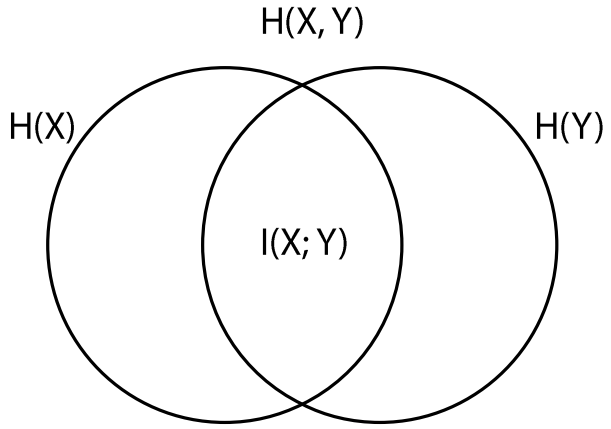

Figure 2 – Information decomposition of a system of two variables

The amount of information that is contributed to the system by each variable separately is quantified by its entropy $H(X)$ and $H(Y)$ respectively. Finally, the shared part of information in the system is quantified by the mutual information of those variables $I(X;Y)$.

$$I(X;Y) = H(X) + H(Y) - H(X,Y) \tag{3.4}$$

The definition of mutual information follows the logic of summing both independent contributions which together include the mutual information twice. Thus, by subtraction of the whole system, it is exactly the mutual information that remains.

However, this decomposition captures only the total information in the system. In order for the system of two variables to only quantify information about a single specific variable S, we need to replace the entropies with mutual information with respect to S. Now, the total information about S in the system is quantified by the mutual information $I(X,Y;S)$ and consistently, information contributed separately by each variable is quantified by $I(X;S)$ and $I(Y;S)$. Following the same construction as in the general

case, the shared part of information between the variables is quantified by $Red_S(X;Y)$ (unconditional version of (eq. 3.2))

$$(3.5)\quad Red_S(X;Y) = I(X;S) + I(Y;S) - I(X,Y;S)$$

However, it can be shown that unlike mutual information, which is a non-negative quantity, Red_S as defined here can be negative (Williams & Beer, 2010). Hence, the total amount of information in the system about S can be larger than the sum of both individual contributions, despite the fact that the shared part of information is included in the sum twice. Therefore, there must be an emergent information when considering both variables together that is not present when they are considered separately. This phenomenon is called synergy and the emergent information is called synergistic information. According to these definitions, DFI corresponds to the difference between the redundancy (the shared part between two variables) and the synergy between X and Y. Hence, DFI is not satisfactory measure for representation of the fourth predicate since it clearly does not represent solely information that is shared between X and Y.

In 2010, Williams and Beer introduced a new framework for information decomposition of mutual information between a system of predictors and a target variable (Williams & Beer, 2010). It decomposes the mutual information into nonnegative terms that, including other beneficial properties, differentiate between redundant and synergistic contributions. This mathematical formalism is also known as Partial Information Decomposition, or PID. To facilitate understanding, we first elaborate on the mutual information between two predictors and a single target variable (more complex system follows). In our case, all variables are represented by time series. Williams and Beer distinguished three types of contribution (Fig. 3) to the information contained by two variables - predictors (here called A and B for distinction

between the formalisms) with respect to the information about a third one - target (C). If A and B both carry the same information about C, and therefore it suffices to know only one of them to fully determine the mutual information between them and C, the information they both carry is called *redundant* (Fig. 3A). If A carries information about C that B does not, the information in A about C is called *unique* (Fig. 3B) and it holds for B symmetrically. Finally, if neither of the variables provides information about C separately, but they do provide information about C when observed together, it is called *synergistic* information (Fig. 3C). An exemplar case of emergent synergistic information is when we assume that all the variables are binary, probability of each predictor having value 0 or 1 is the same $p(A=0) = p(A=1) = 0.5$, equivalently for B, and $C = \mathrm{XOR}(A, B)$. XOR here stands for the exclusive OR operation. In this example, the knowledge of any of the predictors separately does not provide any information about the target variable C as it retains equal chance of being 0 and 1. Only the knowledge of both A and B together determines the value of C.

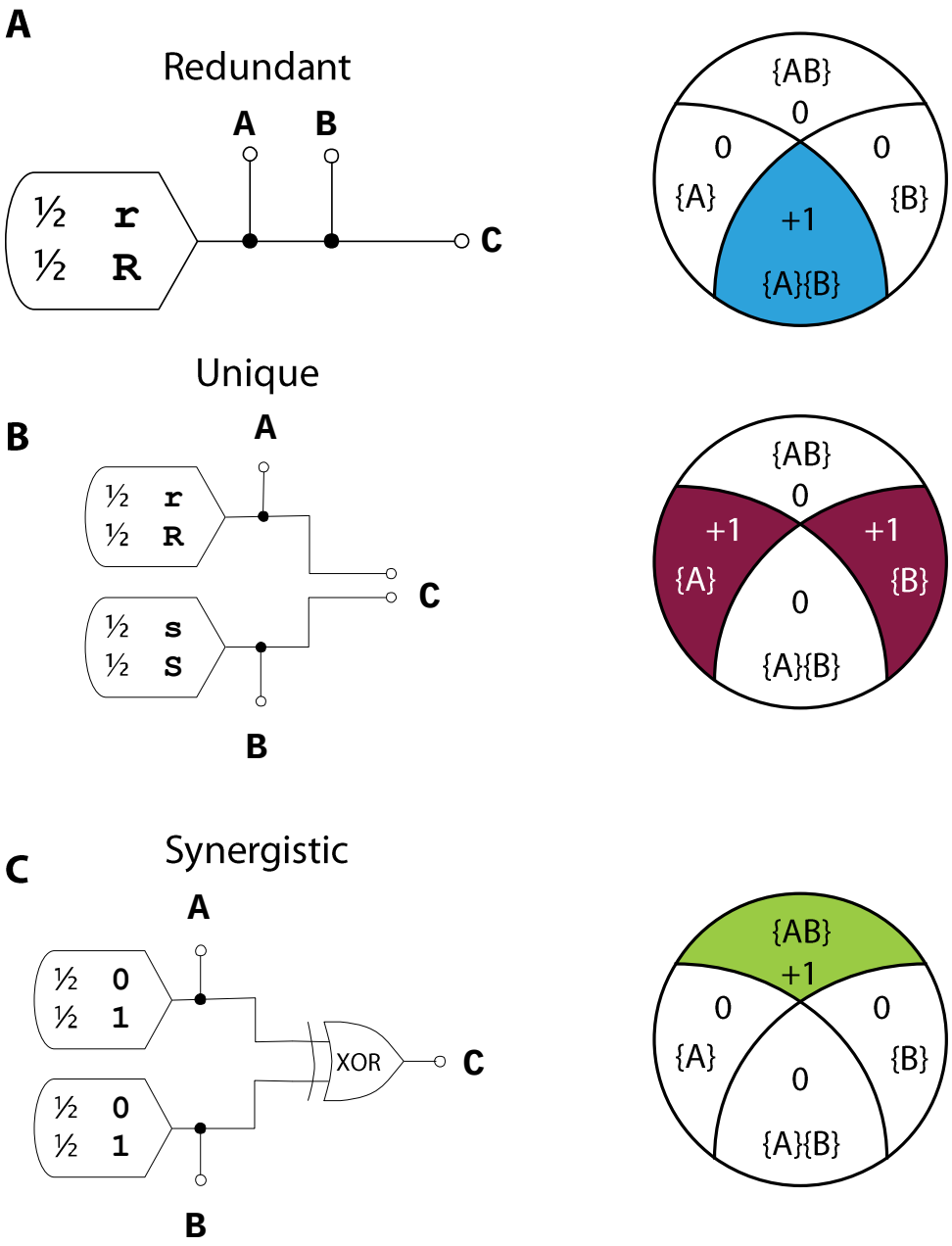

Figure 3 – Visualization of all possible information interactions of two variables. The left column shows a wire diagram that represent the scenario. The source can contribute either a letter or a digit and the number in front of them determines with which probability it occurs. The right column determines the information decomposition of the contribution of both sources. In agreement with (Williams & Beer, 2010) we use notation $\{A\}$ *or* $\{B\}$ *for unique information present in* A *and* B *respectively (the side fields in the decomposition diagrams),* $\{A\}\{B\}$ *for redundancy between them (the bottom field) and* $\{AB\}$ *for their synergy (the upper field).* ***A*** *represents a redundancy where both sources contribute exactly 1 bit of the same information (because the chances for both lower and upper case letter are 0.5 and the sources are linked to each other) to* C*. In* ***B***

the final outcome is composed of two letters that can both be either upper case or lower case and each source determines case of a different letter. Therefore, each source contributes 1 bit of unique information to the outcome. Finally, ***C*** *depicts a synergistic scenario where both sources contribute 1 bit of synergistic information because only the knowledge of both at the same time can determine the outcome of variable* C *. The knowledge of any single one of them does not determine the outcome nor it changes the probability distribution of its possible values. The figure was adapted from (Griffith & Koch, 2014).*

Note that in order to compute the different contributions to a mutual information, it is enough to be able to compute either the synergy or the redundancy because the rest can be then reconstructed from the formalism that was known before. E.g. if we can compute the redundancy between A and B, then the unique information contributed by a single source can be obtained by subtracting the redundancy from the mutual information between the particular source and the target variable $unique(A) = I(A;C) - redundancy(A,B)$. The mutual information $I(A;C)$ cannot contain any synergy as the source A is observed independently on B here. The synergy then comes trivially as it is the only remaining part of the whole mutual information $I(A,B;C)$ after subtraction of all the other parts. The formalism of William and Beer builds on a measure of redundancy called I_{min} which is defined as:

$$(3.6)\quad I_{min}(C;\{A_1,A_2,...,A_k\}) = \sum_c p(c) \min_{A_i} I(C=c;A_i)$$

Where $I(C=c;A_i)$ is a quantity called specific surprise and it is defined followingly:

$$(3.7)\quad I(C=c;A) = \sum_a p(a \mid c)\left[\log\frac{1}{p(c)} - \log\frac{1}{p(c \mid a)}\right]$$

Where a is one of all possible values of A. I_{min} quantifies redundancy in the form of information that is common to all the sources A_i. To continue with the quantification of the decomposition we will consider unique information to be a special case of redundancy (redundancy of a source with itself). This

will allow us to use the redundancy measure to quantify all information in the system. Indeed, it can be shown that in this case I_{min} is equal to a regular mutual information between the source and the target. Note that I_{min} does not quantify the unique contribution of a single part of the decomposition but all redundant information among the sources it is given which can include other parts of the decomposition. For example, as mentioned earlier, $I_{min}(C;\{A\})$ is equal to the mutual information between C and A which, as we showed earlier as well, includes the redundant information shared by A and B. However, as also shown above, it is possible to reconstruct the information contributed independently by all parts of the decomposition by subtracting all the other parts that are included in them. This imposes a natural ordering among all possible parts of the decomposition. All parts necessary for computation of an information contribution of another part are preceding it. Following the logic of the example above, the redundancy between A and B would precede the unique contribution of either A or B because its knowledge is necessary in order to compute the unique contributions.

In order to distinguish between different parts of the decomposition, we will use this notation: redundancy between sources A and B will be written as $\{A\}\{B\}$, unique information of a source A will be $\{A\}$ and synergy between A and B will be noted as $\{A,B\}$. We will call the set of all sources R and the set of all parts of the decomposition based on those sources $\mathcal{A}(R)$. Note that since we will be only computing I_{min} for all elements of $\mathcal{A}(R)$, we only need to include such parts of the decomposition that none is a superset of another as I_{min} i.e. redundancy of such part would be equal to the information contained in the smaller set as adding more sources cannot increase the total redundancy. Also, to avoid confusion, note that $\{A\}\{B\}$ is not a superset of $\{A\}$ because one denotes redundancy and the other one

unique information and the proper notation for redundancy would be $\{\{A\}\{B\}\}$. However, this rather cumbersome notation is not necessary.

Next, we can formally define the ordering on $\mathcal{A}(R)$:

$$(3.8)\ \forall \alpha, \beta \in \mathcal{A}(R), (\alpha \preceq \beta \Leftrightarrow \forall B \in \beta, \exists A \in \alpha, A \subseteq B)$$

This ordering creates a lattice from the elements of $\mathcal{A}(R)$, in which a higher element provides at least as much redundant information as any lower one and the highest element provides all information that is in the mutual information between all sources and the target $I(R;S)$. For example, we show the decomposition of the mutual information between two sources A and B and a target C mentioned above. It only includes the terms that we already discussed: at the very bottom, it is the redundancy between sources, above that their unique contributions and above all there is the synergy (Fig. 4).

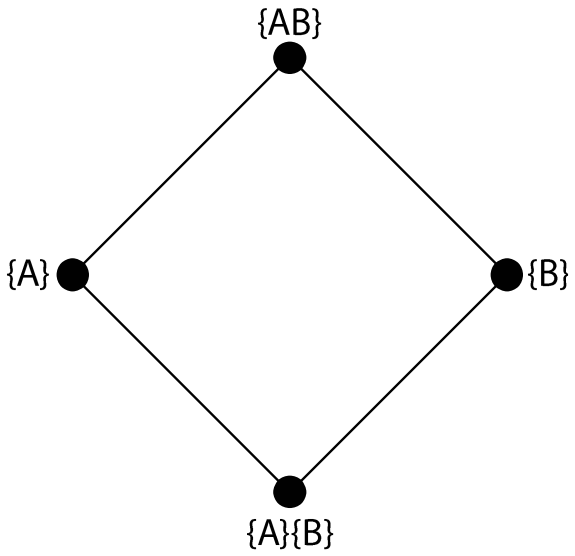

Figure 4 – Partial information decomposition lattice for 2 predictors.

In order to compute the separate contribution of a node in the lattice we need to subtract all information that is provided by nodes below it from the value of I_{min} of the particular node. This amount of information, provided individually by a single node, is called Π_R (partial information term) and it is defined as:

$$\Pi_R(S;\alpha) = I_{min}(S;\alpha) - \sum_{\beta \prec \alpha} \Pi_R(S;\beta) \tag{3.9}$$

Now, it is finally possible to quantify all the individual contributions of all parts of the decomposition.

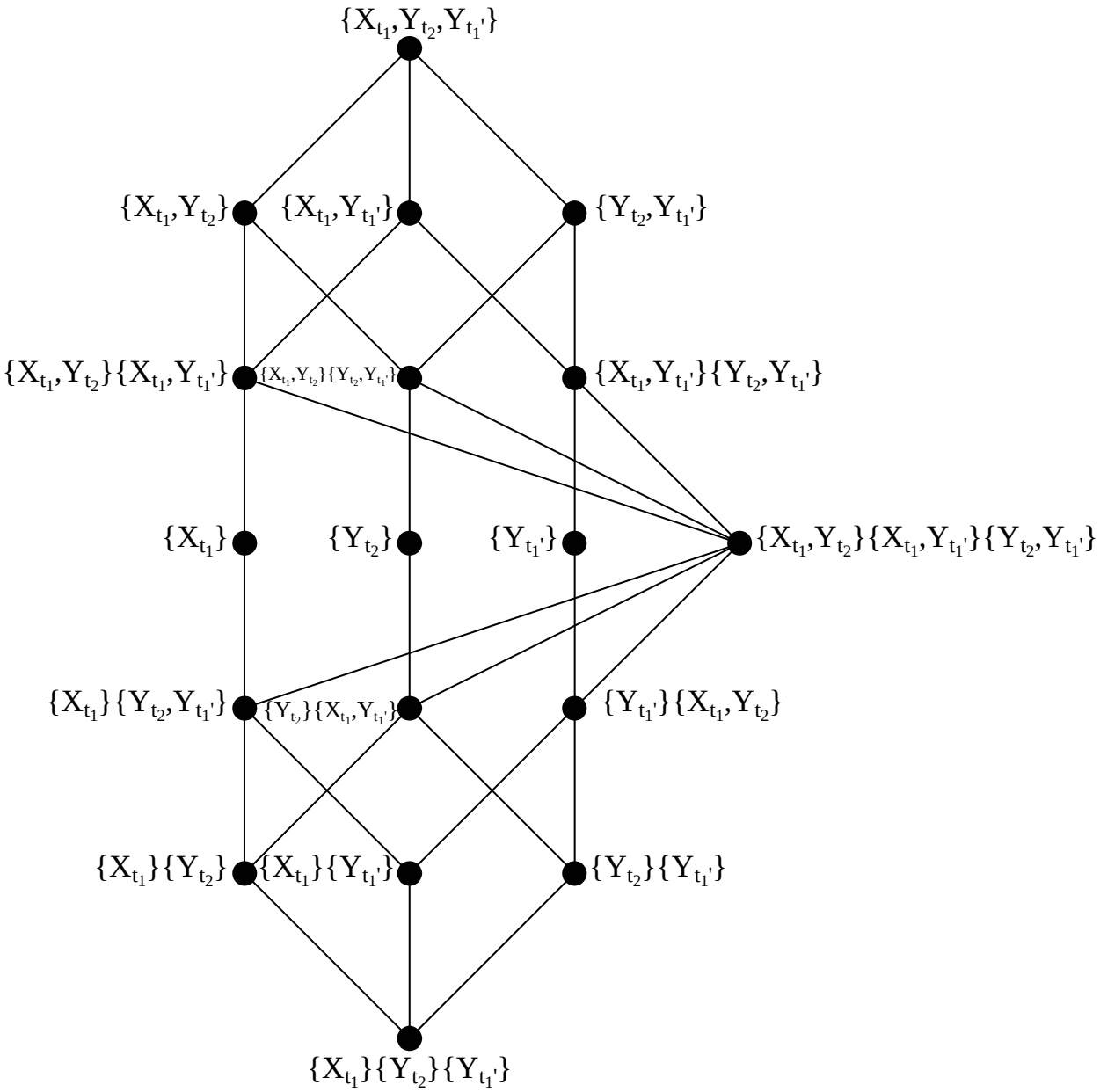

Figure 5 – Partial information decomposition lattice for 3 predictors. This is the lattice showing relations between all partial information terms (Π_R) that arise from the decomposition of information interaction between 3 variables (X_{t_1}, Y_{t_1} and Y_{t_2}) and a target S. The lattice is ordered based on the following: A collection of sources A is considered to "succeed" a collection B (i.e., to be above in the lattice) if for each source in A there exists a source in B that does not provide any additional information than the one in A. The bottom term represents the joint redundancy between all three variables and the top their synergy.

To define the information transfer measure, we consider the decomposition of a system with a set of three predictors $R = \{X_{t_1}, Y_{t_1'}, Y_{t_2}\}$ (Fig. 5) and its mutual information with the target stimulus feature S. To satisfy the aforementioned criteria for such transfer, we remind the reader that we have to quantify the information about S present in Y_{t_2} that is redundant with respect to information already present in X_{t_1} and unique with respect to Y_{t_1}. Consequently, the desired information is expressed by the following partial information term:

$$(3.10)\quad rDFI = \Pi_R(S; \{X_{t_1}\}\{Y_{t_2}\})$$

Note that it is not necessary to apply any conditioning on the past of Y, because this is implicitly accomplished by excluding the information captured by other Π_R terms of the lattice that include the past of Y. As a side note, even though the lattice approach has been generally accepted as correct, the exact measure used to calculate the redundant information (I_{min}) has been criticized for potentially miscategorising purely unique information as synergistic and there were several attempts to improve it (Bertschinger, Rauh, Olbrich, Jost, & Ay, 2014; Griffith & Koch, 2014; Harder, Salge, & Polani, 2013). In the case of only two predictors, the measure of Bertschinger (Bertschinger et al., 2014) is currently preferred, but it cannot be applied for a larger set of predictors than 2. Given that the rDFI involves three predictors and we only use it to express redundancy, we maintain the lattice approach with the original I_{min} of Williams and Beer.

CHAPTER 4: TESTING THE NEW MEASURE OF TRANSMITTED STIMULUS-SPECIFIC INFORMATION WITH BOTH SIMULATED AND REAL DATA

In the previous chapter, we developed a new mathematical measure that is designed to be applied to simultaneous recordings of neural activity from multiple locations and reveal of what information is being transmitted between these neural populations. In this chapter, we test this measure extensively both with simulated data that implement different scenarios of information transmission, and with both real EEG and real MEG data recorded from human subjects. In what follows, we first describe the methods and then the results of these extensive tests.

4.1 METHODS

4.1.1 Poisson Simulations

To test the reliability of our new measure, we performed numerical simulations. We analyzed different scenarios, each representing a particular set of conditions under which stimulus-dependent communication occurs. The structure common to all scenarios comprises a stimulus-dependent transfer of information between two nodes, here called X for the sender of the information and Y for the receiver. Each node carried a stimulus-dependent signal plus a noise drawn from a Poisson distribution with the mean equal to the signal value at a given time point. The presence of a stimulus was simulated by adding a Gaussian profile with duration *ls* and which amplitude was modulated by the stimulus value, to the otherwise constant signal of X at a time t_1. Communication was represented by adding the simulated activity of X in the time interval of the added Gaussian profile to the signal of Y at t_2. The difference between t_1 and t_2 represented the communication delay. Stimulus S values, one per trial, were drawn

from a uniform distribution and the described simulation was independently repeated for every trial. We used 4 possible values of stimulus.

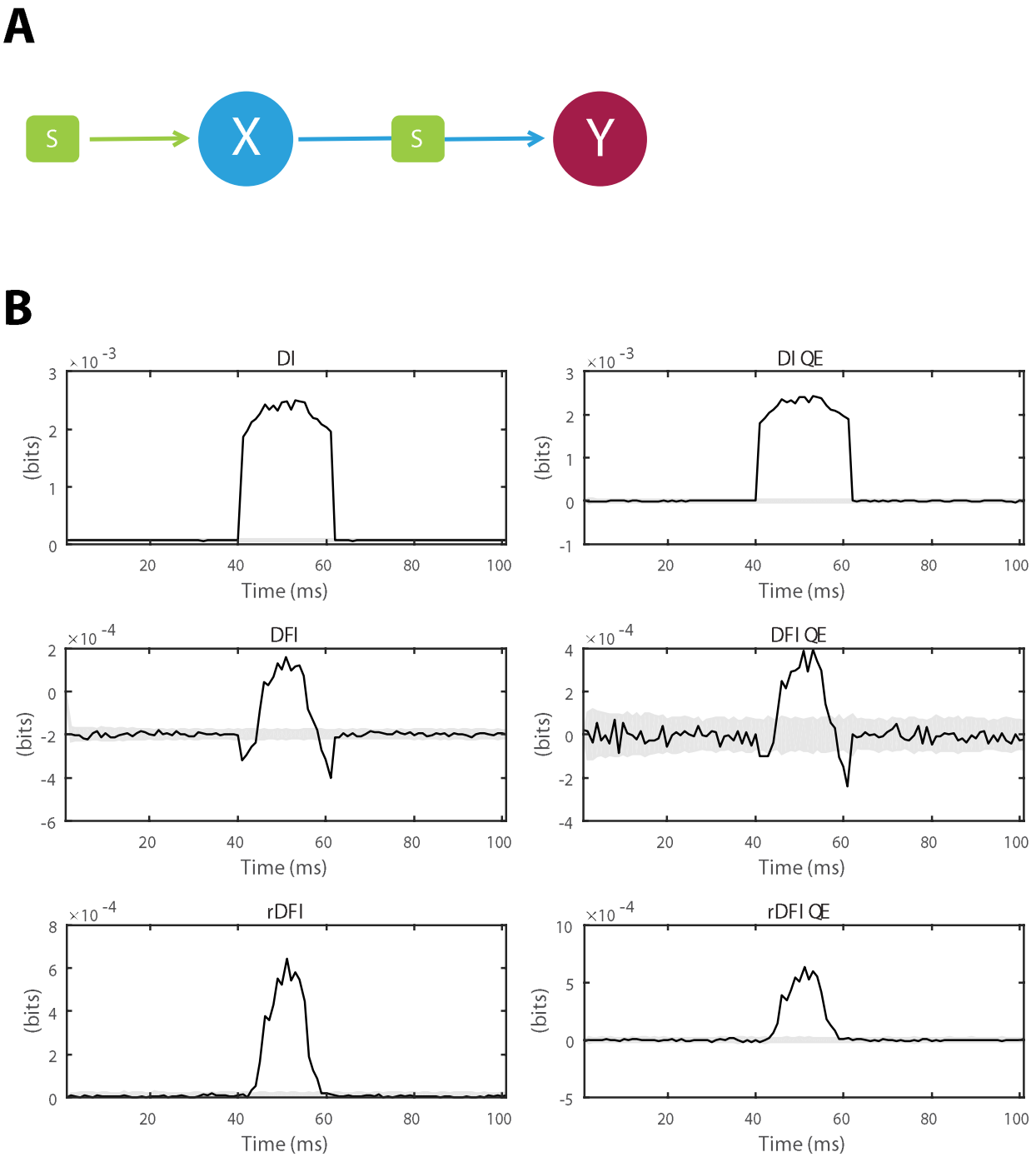

*Figure 6 - **A** Scheme of a simulation scenario representing the most basic transfer of information about S from X to Y without any extra confounders. **B** All information quantities clearly demonstrate a significant peak,*

where the communication occurs, in their both uncorrected and bias corrected (QE) versions. It is noteworthy however, that peaks of DFI are much smaller in comparison to its surrogates (grey area) than those of rDFI. Moreover, DFI uncorrected values outside of the transfer are negative and there are significant dips before and after the transfer peak that are difficult to interpret.

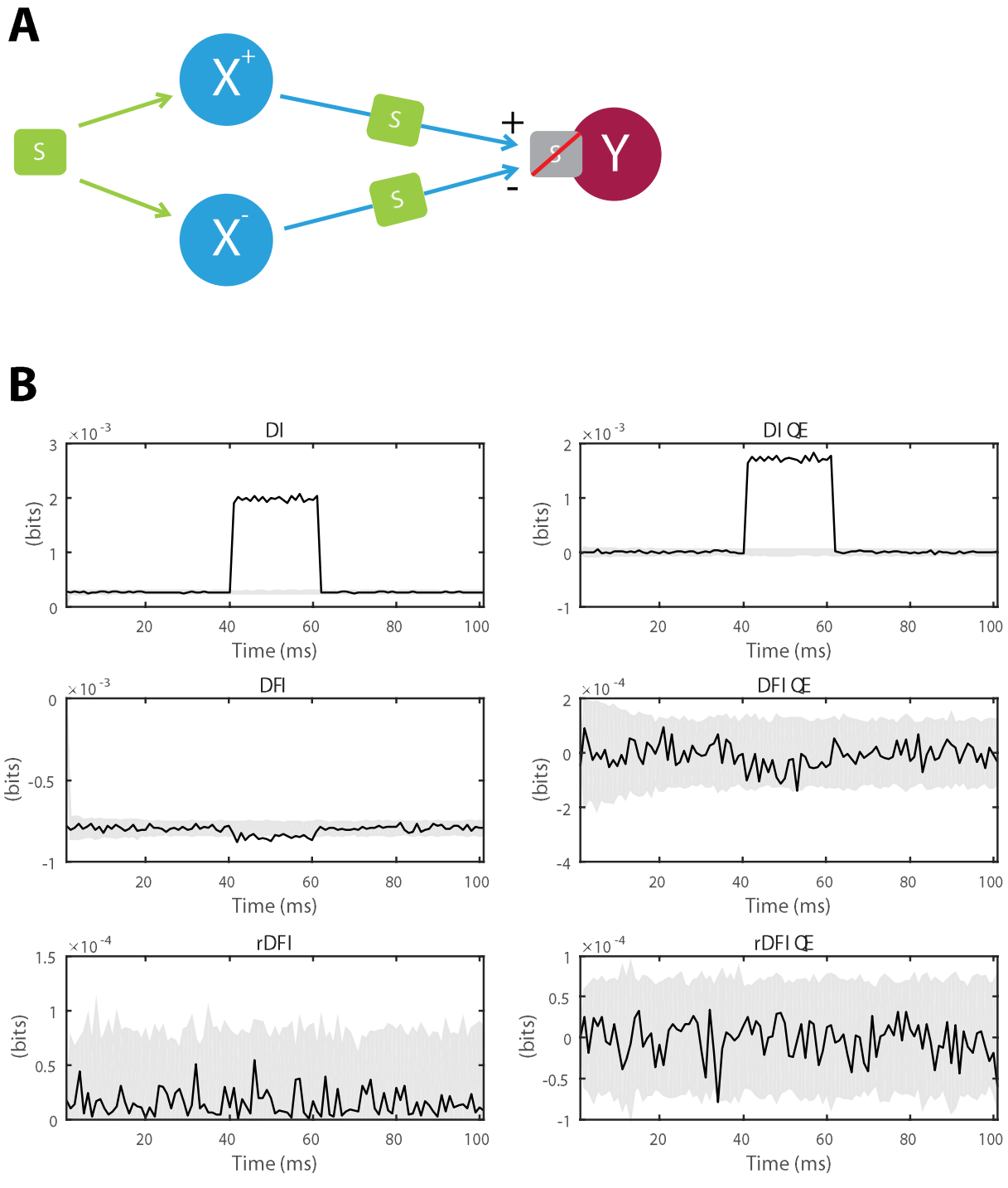

*Figure 7 - **A** Scheme of a simulation scenario representing an information transfer from X to Y that does not convey information about S, despite the information being present in X. The simulation's intent was to confirm whether the measures can distinguish which information is being transferred. **B** First, DI, both corrected and uncorrected, correctly confirms that there is an information transfer between X and Y. On the contrary, neither DFI nor rDFI show a significant peak, which is in line with the design of the scenario. However, DFI, exhibits significant negative values aligned with the transfer.*

All simulations were based on one of the three underlying scenarios. The first scenario (Fig. 6A) simulated a transfer of stimulus-dependent information from a node X to a node Y. The second scenario (Fig. 7A) represented a situation where only stimulus independent information was transferred from X to Y. The aim was to test whether the novel metric efficiently dissociated between the stimulus-dependent and the stimulus-independent transfer. In order to simulate such transfer, the sender node X was split into two sub-nodes that each had its own simulated activity. First, X^+ was simulated as described above and the second one, X^-, had its signal equal to a baseline minus the signal of X^+. We set the baseline signal of X^+ to such a value that the constant part of the signal, where the stimulus representing Gaussian profile was not added, had the same value for both X^+ and X^-. The activity of X^- was computed from its signal identically as in the generic case. To simulate the transfer, activities of both sub nodes were added together and normalized while being added to the signal of the receiver. Note that for computation of the information quantities, the activity of X was represented by a tuple composed of both values of activity of X^+ and X^- at a given time point. Last, the third scenario (Fig. 8A) represented a situation where only stimulus independent information was transferred from X to Y, equally to the scenario above, with added confounding information about a stimulus S' that was drawn from the same probability distribution as S but it was an independent draw. The goal was to test additional confounding effects due to external sources

of information about S. In order to simulate such situation, we made an independent draw of stimulus values S' from the same distribution as for the original ones. Based on those values we added a Gaussian profile, with an equal length as the one described above, to the signal of Y at t_2 that was modulated by the new stimulus values S'.

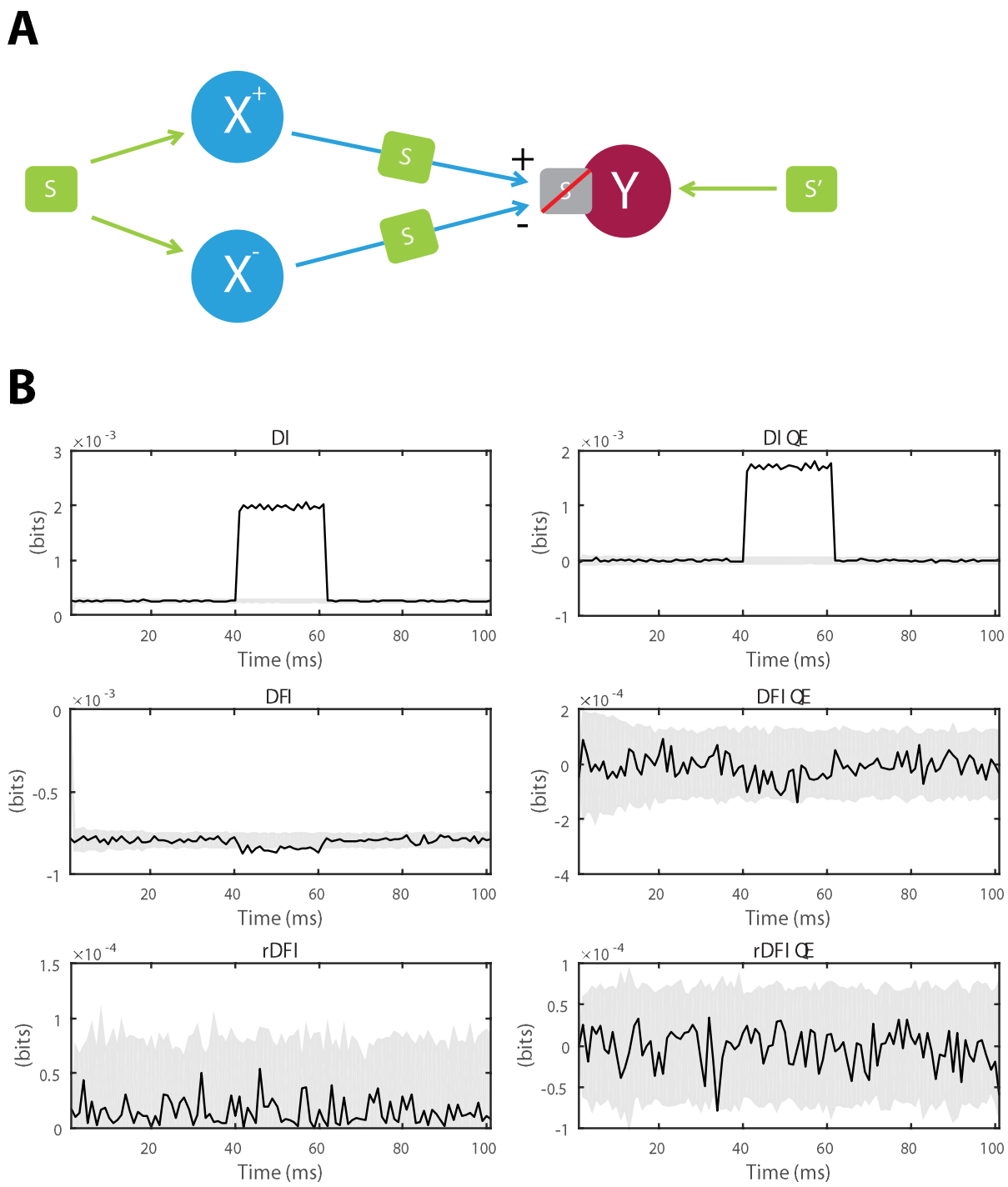

*Figure 8 - **A** Scheme of a simulation scenario representing an information transfer from X to Y that does not convey information about S, though the information is present in X. And an extra source of information about a different stimulus S' drawn from the same probability distribution as S. The simulation's intent was to confirm whether the measures can be confounded by a stimulus with the same probability distribution as S that isn't the same draw as S. **B** In agreement with Fig. 6B, DI, both corrected and uncorrected, correctly*

confirms that there is an information transfer between X and Y. On the contrary, neither DFI nor rDFI show a significant peak, which is also in line with the design of the scenario. However, DFI, as in Fig. 6B, exhibits significant negative values aligned with the transfer.

We conducted three simulations based on the three scenarios, one for each, as described above. Additionally, we also performed four more simulations to test variations of behavior of the rDFI that are not based on the differences in connectivity of the scenarios and therefore, we based them all on the first scenario. In these simulations, we only used 2 stimulus values.

The first of these additional simulations was designed to test whether the negative values of DFI were due to synergistic effects. In this particular simulation, and also in the three described above, we added an extra confounding signal to the signal of both X at t_1 and Y at t_2. This confounding signal was modulated by a random variable drawn from the same distribution as the stimulus, representing a transfer about a different stimulus with similar properties. The signal was added as a Gaussian profile for the three simulations representing different scenarios, but it was in a form of a constant function in case of the high synergy simulation designed to determine the relation between high synergy and negative values of DFI.

Next, we designed a simulation to clarify the effects of conditioning on the past. It was created based on the first scenario, but had varying signal peak width (80ms and 40ms) and varying distance between the peaks (5ms and 20ms). Next simulation was constructed to test directionality sensitivity of the information theoretical measures. It was also based on the first scenario and had signal peaks with width of 80ms and distances between peaks 5ms, 15ms and 25ms. Last, we created a simulation to test effects of noise on the behavior of the measures. It was also based on the first scenario and used the same signal peak widths. However, for high noise condition, we added a sine wave, that had a period of ~200ms and a random phase drawn from a uniform distribution $[0, 2\pi)$ for every trial, to signal of both X and Y. The low noise

condition only had a constant value added to both signals, decreasing the relative difference between the baseline and the peak.

Information quantities on simulated data

For every simulation, we computed values of DI, DFI and rDFI for each time point 1 to 100. These information theoretic quantities require the estimate of the full joint probability distribution. To compute it, we binned all variables for a given time point across trials into 3 bins, using a technique that attempted to distribute equally the number of samples per bin and resulted in the closes possible binning to an equal one. The number of bins was set to 3, as a trade-off between the size of the joint probability distribution, needed for computation of the rDFI, and its computation cost. For the simulations demonstrating robustness to noise, effects of conditioning and high synergy, at each time point t we took all past values of X and Y from a time point $t-d$, where d stands for a delay. The delay was set to the distance between peaks of the stimulus related activity. However, in the scenario demonstrating the effects of conditioning, the previous is true only for the sender X as for the past of Y we computed all information measures for d = 10, 30 and 50 time points.

For the first three simulations, we computed the information quantities at each time point for a set of delays in interval (from 1 to 45 samples) and then took the average of those values as the resulting value. We skipped the first 20 time points as there would not have been enough of possible delays to average over and obtain a reliable value. Therefore, these were computed for time points from 21 to 120 (mapped to 1 to 100 in all figures).

In order to correct for biases due to the limited number of samples, we used the Quadratic Extrapolation procedure (Panzeri, Senatore, Montemurro, & Petersen, 2007). The bias grows quadratically

as a function of the logarithmic decrease of the number of trials. Hence, we first computed the value of a given information quantity using all, a half and a quarter of the available simulated trials, and then we fitted a quadratic function (x axis being the logarithm of number of trials) through the obtained values. The constant coefficient of the quadratic function was the corrected value for the information quantity. To increase accuracy in the information theoretic measures, we computed the values for both halves and all quarters and then averaged across them.

Statistical analysis

In order to assess the level of significance of the different information theoretic measures, we used non-parametric permutation techniques. For every time point for which we computed the information quantities, we established its 0.1th – 99.9th percentile interval under the null hypothesis. To obtain the values under the null hypothesis, we randomly shuffled stimulus order across trials for the computation of significance of DFI and rDFI and values of X for the computation of DI. In this manner, we computed 100 surrogate values based on independent shufflings. To establish the desired percentile values, we then fitted a Gaussian distribution through the surrogates for the measures that can be both positive and negative (DFI and all the (quadratic extrapolation) corrected version of all the measures). For the non-negative measures (uncorrected DI and rDFI) we fitted a Gamma distribution through the surrogates. We performed Chi-Square Goodness of Fit tests on all the fitted distributions. There were no more than 5 time points for which the fit was rejected for any of the measures.

4.1.2 Leaky integrate and fire neuron simulations

In order to generate data qualitatively similar to physiological recordings, we decided to use a model based on the work of Mazzoni et. al (Mazzoni, Panzeri, Logothetis, & Brunel). In their work, they presented

a model of a cortical network composed of leaky integrate and fire neurons and they managed to obtain behavior that strongly resembles behavior of primary visual cortex. To construct the intended scenario, we used the model to create two separate networks that had a unidirectional connection from excitatory neurons in one network to both (excitatory and inhibitory) populations of neurons in the other network (Fig. 9B).

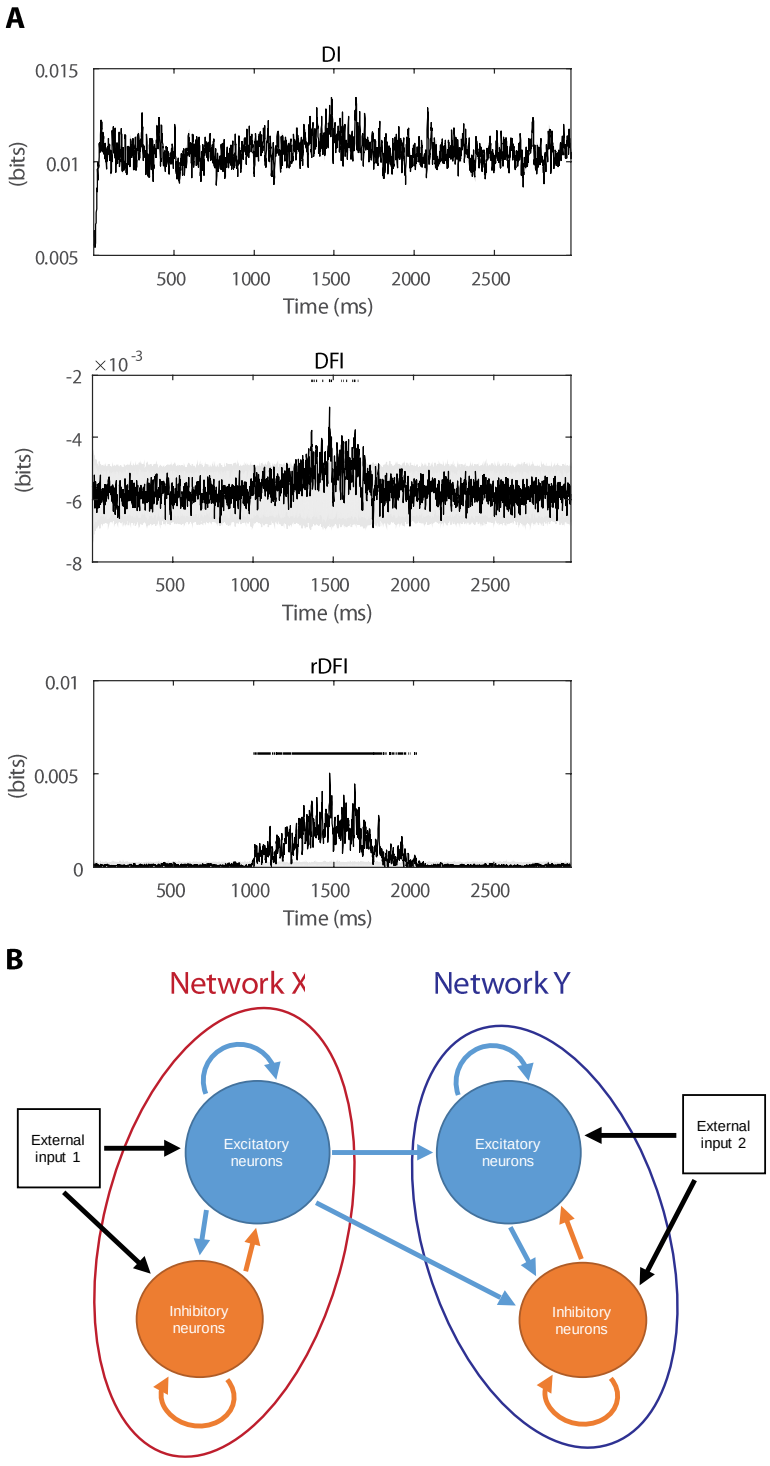

Figure 9 – In **A** *the figure demonstrates similar behaviour as in the basic stimulus transfer scenario, where DI correctly identifies that there is an information transfer present. In contrast to the previous simulations, it is present all the time as the networks are connected throughout the whole simulation. DFI creates a little peak during the time of the stimulus transfer but except of the middle of it, it is not significant (Values with black*

dot/line over them). Finally, rDFI correctly captures the transfer and exhibits a significant peak across the whole time of the transfer. In **B** *we show how the model was constructed. We created two separate networks based on the simulations described in (Mazzoni et al., 2008) and added unidirectional connections from excitatory neurons of one to both populations of the other.*

Each of the simulated networks is composed of $N = 5000$ neurons. 80% of the neurons are taken to be excitatory, the remaining 20% are inhibitory (Braitenberg, 1991). The network is randomly connected: the connection probability between any directed pair of cells is 0.2 (Holmgren, Harkany, Svennenfors, & Zilberter, 2003; Sjöström, Turrigiano, & Nelson, 2001). In case of an inter-network directed connection, there is also 0.2 probability of connection between any pair composed of any cell from the receiver network and an excitatory cell from the sender network. Both pyramidal (excitatory) neurons and interneurons (inhibitory) are described by leaky integrate and fire (LIF) dynamics (Tuckwell, 1988). Each neuron *k* is described by its membrane potential V_k that evolves according to

$$(4.1)\quad \tau_m \frac{dV_k}{dt} = -V_k + I_{Ak} - I_{Gk}$$

where τ_m is the membrane time constant (20ms for excitatory neurons, 10ms for inhibitory neurons, (McCormick, Connors, Lighthall, & Prince, 1985), I_{Ak} are the (AMPAtype) excitatory synaptic currents received by neuron k, while I_{Gk} are the (GABA-type) inhibitory currents received by neuron k. Note that in the equation above we took the resting potential to be equal to zero. When the membrane potential crosses the threshold V_{thr} (18 mV above resting potential) the neuron fires, causing the following consequences: a) the neuron potential is reset at a value V_{res} (11 mV above resting potential), b) the neuron cannot fire again for a refractory time τ_{rp} (2ms for excitatory neurons, 1ms for inhibitory neurons).

Synaptic currents are the linear sum of contributions induced by single pre-synaptic spikes, which are described by a difference of exponentials. They can be obtained using auxiliary variables x_{Ak}, x_{Gk}. AMPA and GABA-type currents of neuron k are described by

$$\tau_{dA}\frac{dI_{Ak}}{dt} = -I_{Ak} + x_{Ak} \quad (4.2)$$

$$\tau_{rA}\frac{dx_{Ak}}{dt} = -x_{Ak} + \tau_m\left(J_{k-exc}\sum_{exc}\delta(t - t_{k-exc} - \tau_L) + J_{k-int}\sum_{int}\delta(t - t_{k-int} - \tau_{L-int}) + J_{k-ext}\sum_{ext}\delta(t - t_{k-ext} - \tau_L)\right) \quad (4.3)$$

$$\tau_{dG}\frac{dI_{Gk}}{dt} = -I_{Gk} + x_{Gk} \quad (4.4)$$

$$\tau_{rG}\frac{dx_{Gk}}{dt} = -x_{Gk} + \tau_m\left(J_{k-inh}\sum_{inh}\delta(t - t_{k-inh} - \tau_L)\right) \quad (4.5)$$

where $t_{k-exc/inh/int/ext}$ is the time of the spikes received from excitatory neurons/inhibitory neurons/inter-network exc. neurons (if a connection from another network is present) connected to neuron k, or from external inputs (see below). τ_{dA} (τ_{dG}) and τ_{rA} (τ_{rG}) are respectively the decay and rise time of the AMPA-type (GABA-type) synaptic current. $\tau_L = 1ms$ and $\tau_{L-int} = 3ms$ are latencies of post-synaptic currents for intra- and inter-network connections respectively. $J_{k-exc/inh/int/ext}$ is the efficacy of the connections from excitatory neurons/inhibitory neurons/inter-network exc. neurons/external inputs on the population of neurons to which k belongs.

Each neuron is receiving an external excitatory synaptic input (last term in the r.h.s. of (eq. 4.3). These synapses are activated by random Poisson spike trains, with a time varying rate which is identical for all neurons. This rate is given by

$$(4.6)\quad \nu_{ext}(t) = \left[\nu_{signal}(t) + n(t)\right]_+$$

where $\nu_{signal}(t)$ represents the signal, and $n(t)$ is the noise. $[\cdots]_+$ is a threshold-linear function, $[x]_+ = x$ if $x > 0$, $[x]_+ = 0$ otherwise, to avoid negative rates which could arise due to the noise term. We use constant signal defined by

$$(4.7)\quad \nu_{signal}(t) = \nu_0$$

where ν_0 is a constant rate equal to 2 spikes/ms. The noise represented by $n(t)$ in (eq. 4.6) is generated according to an Ornstein-Uhlenbeck process.

The activity of each network was summarized by generation of simulated local field potential (LFP). To capture in a simple way the fact that pyramidal cells contribute the most to LFP generation the LFPs are modeled as the sum of the absolute values of AMPA and GABA currents ($|I_A| + |I_G|$) on pyramidal cells in every time point of the simulation.

To construct the intended scenario (Fig. 9B) we simulated two networks with the same set of parameters, one receiving input from the other via the inter-network connections. The internal connections of both networks and their external inputs were generated independently. All the parameter values were in agreement with the original work of Mazzoni (Mazzoni et al., 2008) with addition of synaptic efficacies for inter-network connections J_{k-int} that were equal for excitatory and inhibitory neurons and were drawn

from a uniform distribution from the interval $\langle 0, 0.18)$. We ran the simulation for 3000ms, signal recording frequency 1000Hz and the communication signal being present in the sender between 1000ms and 2000ms time points. Due to computational feasibility, we could not run the statistical analysis, described above, for these simulations.

4.1.3 Brain dataset 1: EEG and face detection task

Experimental conditions and behavioral tasks

We tested the novel measures on an EEG dataset publicly available and published by Rousselet et al. in (Rousselet, Ince, van Rijsbergen, & Schyns, 2014). The data were recorded during a face detection task where subjects were presented with an image hidden behind a bubble mask. In a half of the trials the image was a face and the other half contained a random texture. For our purposes, we only considered correct trials where the face was presented (approximately 1000 trials per subject) to the subject (16 subjects). The dataset in the published form is already preprocessed and contains information about which electrode is the left and the right occipito-temporal sensor (the one with the highest mutual information between its signal and visibility of the contralateral eye). It only contains data from 15 subjects as one was discarded for low quality of the recording.

Information theoretic measures

We computed the first derivatives of values of the EEG signal for both occipito-temporal sensors, marked in the dataset, and used both absolute values and the derivatives as a joint signal to compute the information quantities. To establish the joint probability distribution for the computation of the information quantities, we binned both the derivatives and the absolute values into 3 equally populated bins resulting into 9 possible values for each random variable. The visibility of an eye was considered as the

stimulus and it was also binned into 3 bins. We computed the information quantities for all combinations of the direction of the transfer (left to right, right to left) and the particular eye (left and right). These values were computed for every subject and then we averaged across them.

Statistical analyses

We established significant clusters in the time versus delay data grid using a cluster-based nonparametric statistical test introduced in (Maris & Oostenveld, 2007). First, we computed the surrogate values of the information quantities for the whole 2D grid of time and delay as described above, performing 100 permutations. For each of those permutations, we created clusters of values higher than a threshold, based on 8-adjacency in the 2D grid. We set the threshold to 97.5^{th} percentile, that we obtained by ranking all the values in the grid. Then, we computed cluster level statistic by summing all values of a given information quantity within the cluster. Finally, we took the largest cluster-level statistic for each of the permutations and compared them to cluster-level statistic in the non-shuffled data. We considered as significant those clusters which cluster-level statistic was lower than the cluster-level statistic of less than 5 clusters obtained from the shuffled data. Hence, we only considered significant those that would score higher than 95^{th} place in a rank test.

Robustness with respect to the sample size bias

We compared values of (sample size bias) corrected and non-corrected information quantities under two conditions, the information transfer being present in the simulation or not. To compute values for each condition, we determined for which of the areas of the two-dimensional grid of time and delay the transfer was and was not present. The representative value of the first condition was an average across all values in the area with the presence of a transfer and the representative value of the second condition was

an average across values in an identical area moved in time in such manner that it only encapsulated points with no transfer. We computed these values for both conditions (transfer being simulated or not) in simulations that had numbers of trials equal to <2^8, 2^9 … 2^18> and each simulation was repeated 5 times across which we averaged all the computed values.

4.1.4 Brain dataset 2: MEG and visuomotor task

Experimental conditions, behavioral tasks and brain data acquisition

We also analyzed an MEG dataset collected while participants performed an associative visuomotor mapping task, where the relation between a visual stimulus and a motor response is arbitrary and deterministic (Brovelli et al., 2017; Brovelli, Chicharro, Badier, Wang, & Jirsa, 2015). The task required participants to perform a finger movement associated to a digit number: digit "1" instructed the execution of the thumb, "2" for the index finger, "3" for the middle finger and so on. Maximal reaction time was 1s. After a fixed delay of 1 second following the disappearance of the digit number, an outcome image was presented for 1s and informed the subject whether the response was correct, incorrect, or too late (if the reaction time exceeded 1s). Incorrect and late trials were excluded from the analysis, because they were either absent or very rare (i.e., maximum 2 late trials per session). The next trial started after a variable delay ranging from 2 to 3s (randomly drawn from a uniform distribution) with the presentation of another visual stimulus. Each participant performed two sessions of 60 trials each (total of 120 trials). Each session included three digits randomly presented in blocks of three trials. The average reaction time was 0.504s ± 0.004s (mean ± s.e.m.).

Anatomical T1-weighted MRI images were acquired for all participants using a 3-T whole-body imager equipped with a circular polarized head coil. Magnetoencephalographic (MEG) recordings were

performed using a 248 magnetometers system (4D Neuroimaging magnes 3600). Visual stimuli were projected using a video projection and motor responses were acquired using a LUMItouch® optical response keypad with five keys. Presentation® software was used for stimulus delivery and experimental control during MEG acquisition.

MarsAtlas

Single-subject cortical parcellation was performed using the *MarsAtlas* brain scheme (Auzias, Coulon, & Brovelli, 2016). After denoising using a non-local means approach (Coupé et al., 2008), T1-weighted MR-images were segmented using the FreeSurfer "recon-all" pipeline (http://freesurfer.net). Grey and white matter segmentations of each hemisphere were imported into the BrainVisa software and processed using the Morphologist pipeline procedure (http://brainvisa.info). White matter and pial surfaces were reconstructed and triangulated, and all sulci were detected and labeled automatically (Mangin et al., 2004; Perrot, Rivière, & Mangin, 2011). A parameterization of each hemisphere white matter mesh was performed using the Cortical Surface toolbox (http://www.meca-brain.org/softwares/). It resulted in a 2D orthogonal system defined on the white matter mesh, constrained by a set of primary and secondary sulci (Auzias et al., 2013). The parcels corresponding to the subcortical structures were extracted using Freesurfer (Fischl et al., 2002). The subcortical structures included in the brain parcellation were the caudate nucleus, putamen, nucleus accumbens, globus pallidus, thalamus, amygdala, and hippocampus. The whole-brain parcellation therefore comprised 96 areas (41 cortical and 7 subcortical areas per hemisphere).

Single-trial high-gamma activity (HGA) in *MarsAtlas*

MEG signals were down-sampled to 1 kHz, low-pass filtered to 250 Hz and segmented into epochs aligned on finger movement (i.e., button press). Epoch segmentation was also performed on stimulus onset and the data from -0.5 and -0.1 s prior to stimulus presentation was taken as baseline activity for the calculation of the single-trial high-gamma activity (HGA). Artefact rejection was performed semi-automatically and by visual inspection. For each movement-aligned epoch and channel, the signal variance and z-value were computed over time and taken as relevant metrics for the identification of artefact epochs. All trials with a variance greater than 1.5*10-24 across channels were excluded from further search of artefacts. Metrics such as the z-score, absolute z-score, and range between the minimum and maximum values were also inspected to detect artefacts. Channels and trials displaying outliers were removed. Two MEG sensors were excluded from the analysis for all subjects.

Spectral density estimation was performed using multi-taper method based on discrete prolate spheroidal (slepian) sequences(Mitra & Pesaran, 1999; Percival & Walden, 1993). To extract high-gamma activity from 60 to 120, MEG time series were multiplied by k orthogonal tapers (k = 8) (0.15s in duration and 60Hz of frequency resolution, each stepped every 0.005s), centered at 90Hz and Fourier-transformed. Complex-valued estimates of spectral measures $X^{n}_{sensor}(t,k)$, including cross-spectral density matrices, were computed at the sensor level for each trial n, time t and taper k.

Source analysis requires a physical forward model or leadfield, which describes the electromagnetic relation between sources and MEG sensors. The leadfield combines the geometrical relation of sources (dipoles) and sensors with a model of the conductive medium (i.e., the headmodel). For each participant, we generated a headmodel using a single-shell model constructed from the segmentation of the cortical tissue obtained from individual MRI scans as described in section 3.2 (Nolte, 2003). Leadfields were not

normalized. Sources were placed in the single-subject volumetric parcellation regions. For each region, we computed the number of sources *nSP* as the ratio of the volume and the volume of a sphere of radius equal to 3 mm. The k-means algorithm (Tou & González, 1974) was then used to partition the 3D coordinates of the voxels within a given volumetric region into *nS* clusters. The sources were placed at the center of each partition for each brain region. The headmodel, source locations and the information about MEG sensor position for both models were combined to derive single-participant leadfields. The orientation of cortical sources was set perpendicular to the cortical surface, whereas the orientation for subcortical sources was left unconstrained.

We used adaptive linear spatial filtering (Van Veen, Van Drongelen, Yuchtman, & Suzuki, 1997) to estimate the power at the source level. In particular, we employed the Dynamical Imaging of Coherent Sources (DICS) method, a beamforming algorithm for the tomographic mapping in the frequency domain (Gross et al., 2001), which is a well suited for the study of neural oscillatory responses based on single-trial source estimates of band-limited MEG signals (for a series of review see,(Hansen, Kringelbach, & Salmelin, 2010)). At each source location, DICS employs a spatial filter that passes activity from this location with unit gain while maximally suppressing any other activity. The spatial filters were computed on all trials for each time point and session, and then applied to single-trial MEG data. DICS allows the estimate of complex-value spectral measures at the source level, $X^n_{source}(t,k) = A(t)X^n_{sensor}(t,k)$, where $A(t)$ is the spatial filter that transforms the data from the sensor to source level and $X^n_{sensor}(t,k)$ is the complex-valued estimates of spectral measures, including cross-spectral density matrices, computed at the sensor level for each trial *n*, time *t* and taper *k* (for a detailed description of a similar approach see (Hipp, Engel, & Siegel, 2011)). The single-trial high-gamma power at each source location was estimated by multiplying the complex spectral estimates with their complex conjugate, and averaged over tapers *k*, $P^n_{source}(t) =$

$\langle X^n_{source}(t,k) X^n_{source}(t,k)^* \rangle_k$, where angle brackets refer to the average across tapers and * to the complex conjugate. Single-trial power estimates aligned on movement and stimulus onset were log-transformed to make the data approximate Gaussian and low-pass filtered at 50Hz to reduce noise. Single-trial mean power and standard deviation in a time window from -0.5 and -0.1 s prior to stimulus onset was computed for each source and trial, and used to z-transform single-trial movement-locked power time courses. Similarly, single-trial stimulus-locked power time courses were log-transformed and z-scored with respect to baseline period, so to produce HGAs for the prestimulus period from -1.6 to -0.1 s with respect to stimulation for subsequent functional connectivity analysis. Finally, single-trial HGA for each brain region of *MarsAtlas* was computed as the mean z-transformed power values averaged across all sources within the same region.

Whole-brain information theoretic measures

We computed information theoretic measures to quantify the information transfer about visuomotor processing between all pairs of brain regions. To estimate the join probability distribution for computation of the information quantities at each time stamp, we concatenated the HGA contained in windows of 20 msec (4 time instants), trials and subjects, and we binned the values into 3 equally populated bins. We considered as the "stimulus" dimension the HGA either from the baseline or from the event-related interval (2 bins), thus resulting into 6 possible values for each random variable.

4.2 RESULTS

4.2.1 Simulations

Our aim was to develop a model-free measure that quantifies the information about a stimulus S in a time series X that is later present in a time series Y and is new with respect to the information about S already carried by the past of Y. Here, we propose a novel information theoretic measure, named rDFI (eq. 3.10), that exploits latest advancements in the field of Information Theory, in particular the Partial Information Decomposition framework (Williams & Beer, 2010). We first tested the reliability of the novel measure to capture stimulus-specific information transfer in multiple settings through numerical simulations. We compared the results from rDFI with those of DI and DFI.

We designed three scenarios exemplifying three types of information transfer between X and Y. We ran simulations of those scenarios, where we simulated both X and Y by creating a signal with a peak modulated by the stimulus S at t_1 for X and modulated by activity of X at t_2 for Y and using this signal as a mean value for Poisson random process. Numbers drawn from the Poisson distribution were then taken as the activity of X and Y. Unless stated otherwise, we call the difference between t_1 and t_2 the delay of the past.

The first scenario modeled a feedforward transfer (Fig. 6A). As expected, both DI, DFI and rDFI captured the transfer and showed a significant peak where the transfer was present (Fig. 6B). The second scenario represented a transfer between nodes X and Y that did not contain information about the stimulus S while information about S was present in X and no presence of information about S was in Y (Fig. 7A). This scenario was design to test the ability of all the measures to distinguish the transfer carrying or lacking information about the stimulus S. We observed that the transfer was demonstrated by

a peak in DI while there were no significant positive values for either DFI or rDFI (Fig. 7B). We should note that DFI yielded significant negative values (discussed later). Finally, the third scenario simulated a connectivity layout similar to the second scenario with the addition of information about a different stimulus S' with the same probability distribution as S to Y (Fig. 8A). This scenario was designed to reveal whether the different measures can distinguish between stimuli with the same probability distribution. Information transfer was shown as an increase in DI, and the lack of stimulus-specific transfer was correctly quantified by both the DFI and rDFI (Fig. 8B). Overall, rDFI demonstrated behavior that was always in line with the design of a scenario, even in cases where DFI resulted in significant negative values.

After testing of effects of different connectivity patterns, we measured how other changes in the simulation setup influenced DI, DFI and rDFI. We modulated the width of the information peaks in the signal of X, the distance between those peaks, the delay based on which the past values are determined and added an extra noise to the signal. We performed these tests using the feedforward stimulus transfer scenario from the previous experiment (Fig. 10A). First, we tested how well the measures detect presence of information about S in the past of Y since that is one of the predicates that define the measure.

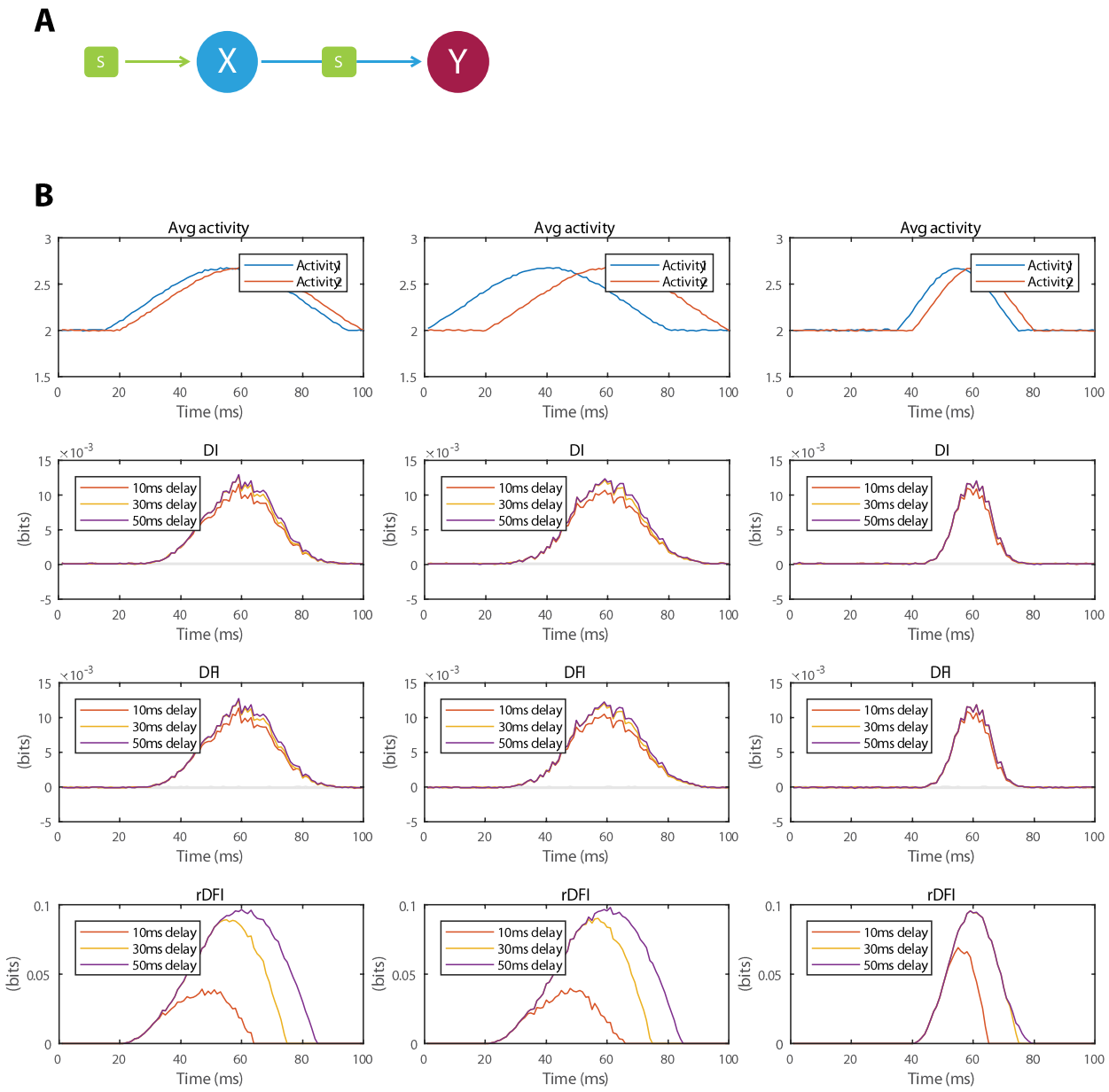

Figure 10 - In order to test influence of other simulation parameters on behavior of the measures, we designed a simulation, using the basic transfer scenario ***A****, that tests influence of width of information peaks, their distance and delay with which we measure the values of past of* X *and past of* Y*. The delay with we measured the past of* X *was fixed to the distance between the peaks, whereas the delay with which we took the values of the past of* Y *varied between 10, 30 and 50ms. In all scenarios the transfer peaks at 60ms. The first column in* ***B*** *depicts wide information peaks in* X *and* Y *with very short distance between them, the second shows same peaks but large distance between them and finally the third column shows narrow peaks with short distance. It can be observed that DI and DFI are not changed by the delay of past, nor by the distance between peaks but their peak width scales with the width of information peaks in* X *and* Y*. On*

contrary, rDFI is not only modulated by the size of the original peaks, but it also changed with the delay of the past of Y. rDFI is lower in general and tends to return to 0 after the peak with decrease in the delay of past. That is in line with the fact that with decrease in the delay we increase the amount of information shared between the past and the present and rDFI correctly recognizes that. However, DI and DFI are not able recognize the presence of information about S in the past of Y in this simulation.

As it can be observed in (Fig. 10B), values of DI and DFI do not change with varying delay of the past of Y nor the distance between information peaks in the signal. They were modulated only by the size of those peaks. Values of rDFI decrease with decreasing delay between the past and the present of Y .These changes appeared independently on any other modulation of the activity (distance between the peaks and their width), confirming that rDFI is able to determine whether the present information in Y was already present at the earlier time point in Y as well. The lower the delay, the earlier values of the past of Y got higher than values of the present of Y. In such case all the information about the stimulus was already present in the past of Y and therefore drove the values of rDFI to 0. The effect is minimal for the delay of 50ms because the past of Y only becomes larger than its present at 65ms of the 80ms peak.

Since there was a discrepancy between rDFI and other measures in the ability of discounting information that was already present in the receiver in the past, we also investigated the ability of the information measures to distinguish the correct directionality of information transfer. We used the previous scenario and measured all quantities in both directions of the transfer even though only the transfer from X to Y was present.

Values of DI and DFI (Fig. 11) for the direction in which there was no transfer present were increasing with decrease in the delay between the information peaks in X and Y, leading to inability to distinguish the direction of the transfer when the delay was too low. rDFI demonstrated consistent behavior

across all simulations, correctly capturing the transfer in the direction in which the transfer occurred and not reaching significant values in the other.

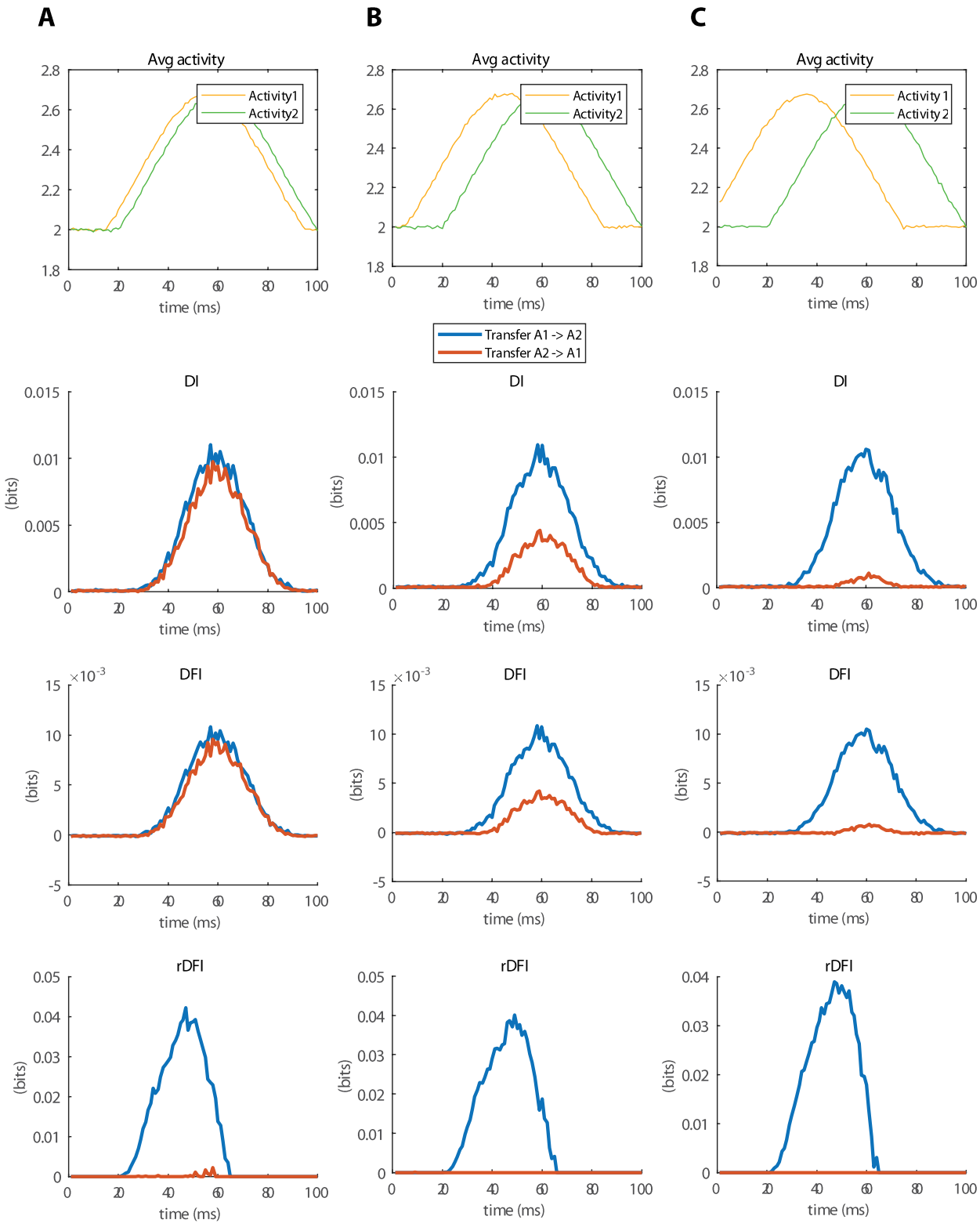

Figure 11 - Based on the previous discovery of DI and DFI not being able to detect presence of information about S in the past of Y, we designed a simulation, using the basic connectivity scenario, to test how the

measures behave when presented with peaks that appear in contradiction with causality. Therefore, we took three values of distance between the peaks, decreasing their overlap and computed all the measures first, using the blue line as X (blue line in inf. quant. values) and conversely, second, using the red line as X (red line in inf. quant. values), thus creating an impossible transfer since the peak in the sender's signal occurs only after the peak in the signal of the receiver. The results clearly show that DI and DFI were not driven to zero if the transfer was against causality and there was a sufficient overlap between the information peaks in time. In comparison, rDFI was significant for the scenario with transfer being in line with causality and remained insignificant for the other.

We tested many simulation settings that could potentially lead to confusion of the rDFI measure. Therefore, we also tested its robustness to general noise in the time series. Again, using the basic scenario (Fig. 10A) and adding either low or high amounts of noise, we established that rDFI is more robust to noise than DFI and DI which can become insignificant along the entire time of the simulation with amounts of noise with which rDFI can still recognize the information transfer.

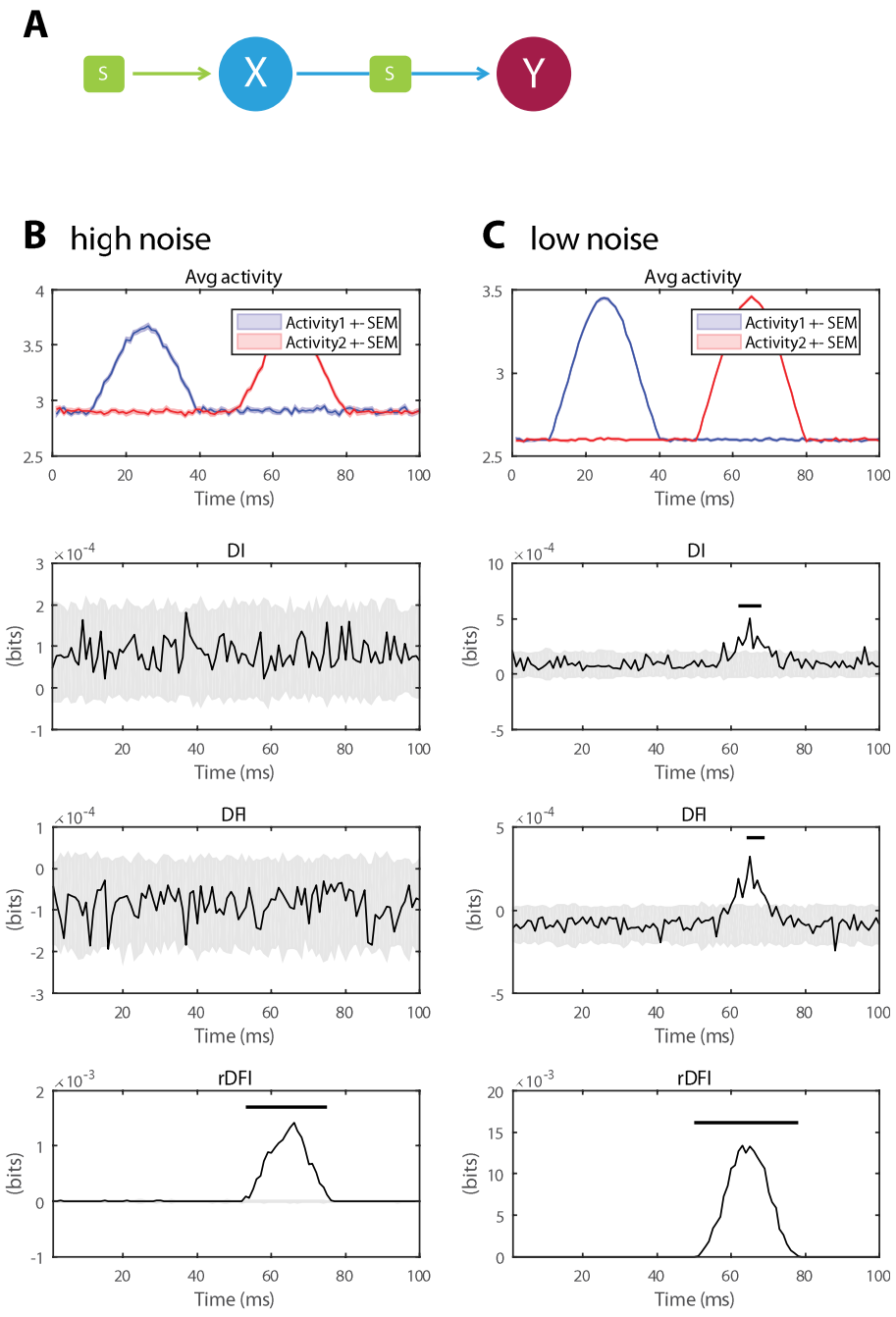

*Figure 12 - This figure shows results of testing of robustness of the measures against noise in the signal. To test that, we used the basic transfer scenario **A** and introduced an extra noise into the signal. The results clearly show that DI and DFI are much more susceptible to noise as their values were insignificant all over the time course in case of the high noise scenario **B** and creating a significant peak with much smaller width than the transfer in case of the low noise scenario **C**. rDFI was able to correctly detect the transfer in both cases, suggesting that it is more robust to noise than DI and DFI.*

During the process of deriving rDFI we also decomposed DFI in terms of the PID framework (Williams & Beer, 2010) (Sup. Material). The decomposition showed that DFI can be seen as a sum of redundancies minus a sum of synergies, suggesting that synergistic effects lead to negative values of the measure. Therefore, we designed a highly synergistic scenario (Fig. 13), where the presence of synergy is confirmed by the co-occurrence of signal similarity and correlations of opposite sign (Pola et al., 2003). In our case, the stimulus dependent correlations had an opposite sign than the signal similarity and they were orders of magnitude higher (in sense of absolute values) than the stimulus independent correlations that had the same sign as the signal similarity.

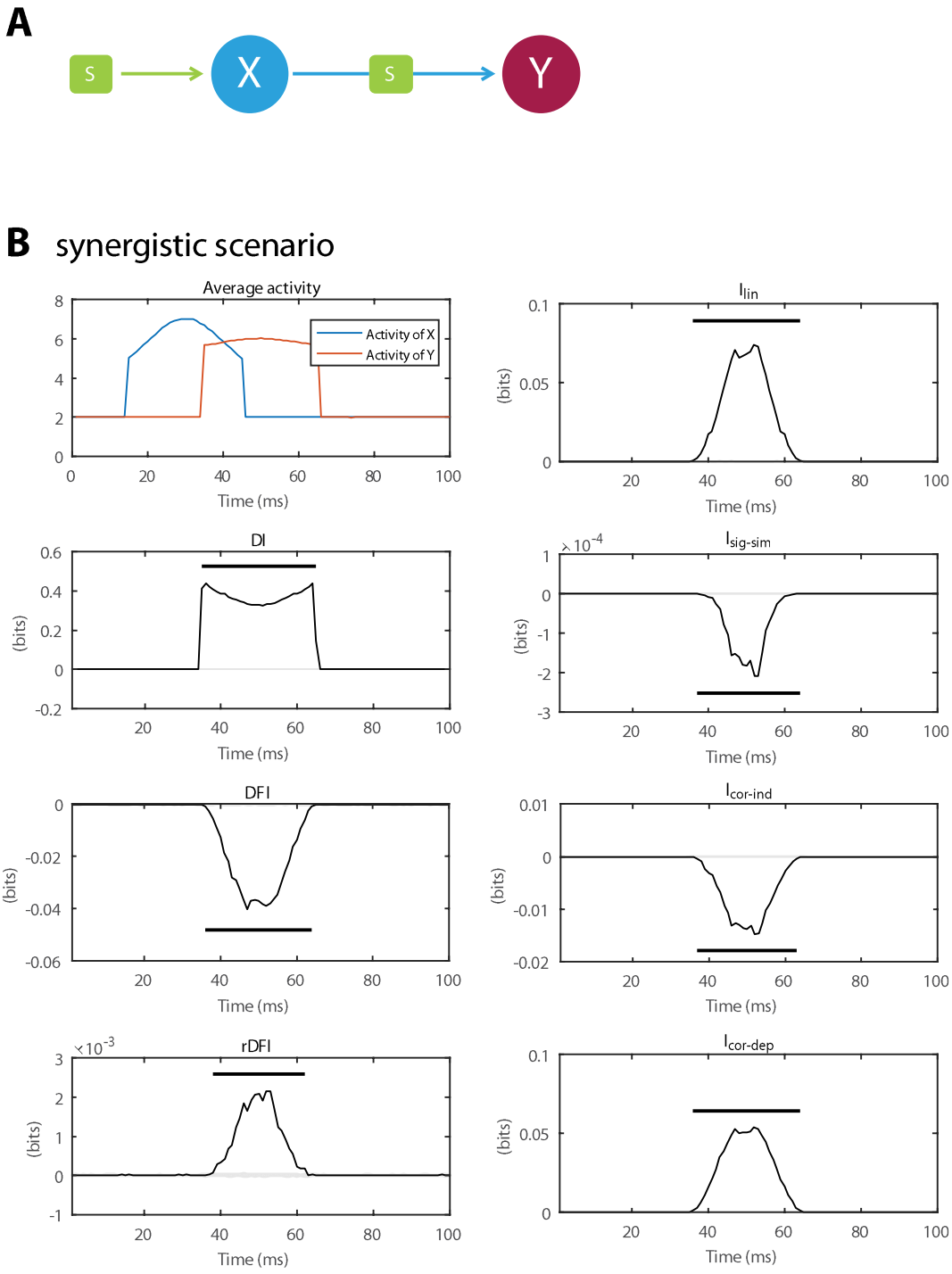

*Figure 13 - To confirm the hypothesis of rDFI being negative because of synergistic effects, we developed a simulation based on the basic transfer scenario **A** that introduced high synergy into the system. **B** This fact was confirmed by presence of the negative signal similarity and the positive stimulus dependent correlations ($I_{cor\text{-}dep}$) (that dominate the stimulus independent correlations ($I_{cor\text{-}ind}$)) as described by (Pola et al., 2003). It can be observed that DFI values created a significant negative peak at the position of the transfer, confirming*

that DFI values unlike DI and rDFI, are decreased by presence of synergy. The top left panel shows the average activity of both nodes X and Y. Other left panels show values of the discussed information transfer measures and the right panel shows all the terms of the information decomposition by (Pola et al., 2003) – linear term (I_{lin}), signal-similarity term ($I_{sig\text{-}sim}$), stimulus-independent correlations term ($I_{cor\text{-}ind}$) and stimulus dependent correlations ($I_{cor\text{-}dep}$).

As hypothesized, DFI yielded a significant negative peak aligned with the presence of the transfer while DI and rDFI demonstrated a positive peak at the same time. It confirms that values of DFI depend on the ratio between redundancy and synergy in the transfer.

Previously, it was shown that information theoretical quantities are susceptible to the sampling bias (Panzeri et al., 2007; Panzeri & Treves, 1996). Since rDFI is based on similar principles, we also tested its behavior on all three scenarios, representing different connectivity, that were mentioned earlier, with varying number of samples (Fig. 14). It confirmed that rDFI values grow quadratically with logarithm of the number of samples which is in line with biases observed in (Panzeri et al., 2007) and that it can also be corrected for by using the quadratic extrapolation procedure, introduced in (Strong, Koberle, van Steveninck, & Bialek, 1998). Using the correction, we only need ≈8 times more trials than the number of unique values of the full joint probability distribution that need to be sampled. The number of those values is equal to the product of the numbers of possible values (numbers of bins) for stimulus, past of X, present of Y and past of Y. For illustration, assuming we used 3 possible values (bins) for each of those quantities, the size of the full joint probability distribution is $3^4 = 81$. Therefore, using the Quadratic Extrapolation method, to obtain reliable values of rDFI we would need ≈650 data samples.

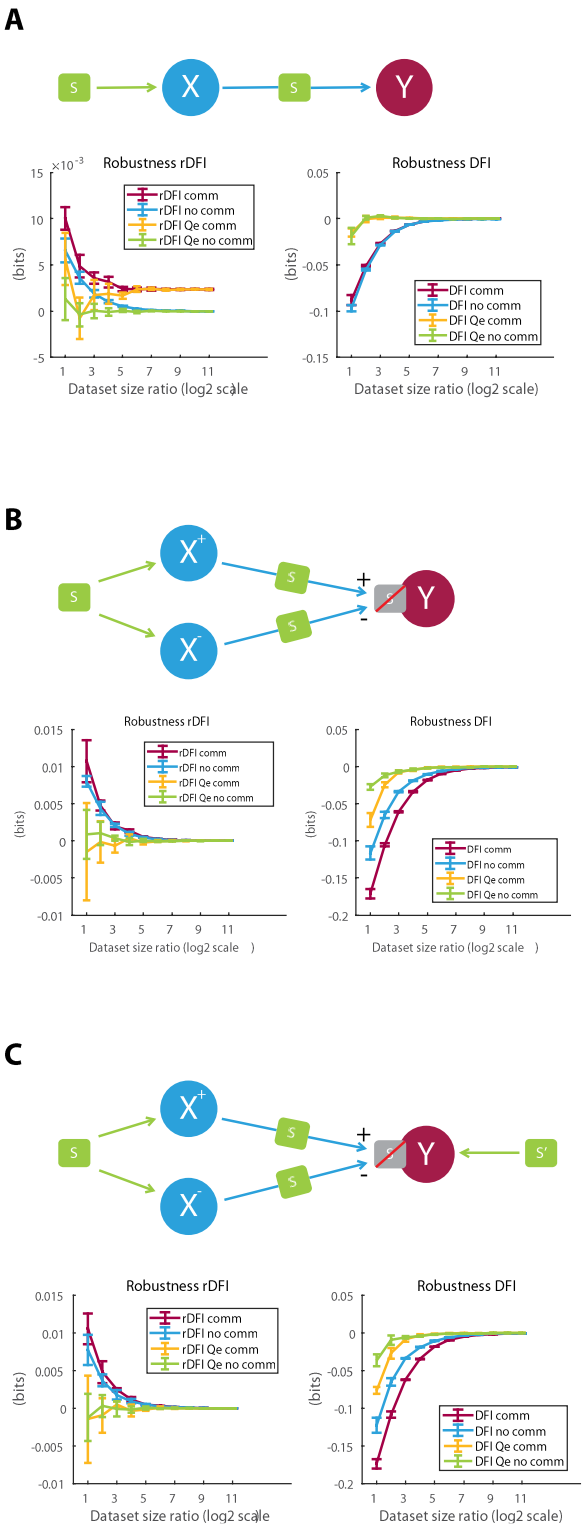

Figure 14 - The figure depicts results of robustness of rDFI and DFI with respect to sampling size bias, which is a common problem for information theoretical quantities (Panzeri et al., 2007; Panzeri & Treves, 1996). We tested the measures in all three connectivity scenarios as described above, for time points where transfer

did (comm) and did not (no comm) occur and compared an uncorrected and corrected computation of them. For correction, we used the Quadratic Extrapolation (QE) correction (Panzeri et al., 2007; Strong et al., 1998). It can be observed that both rDFI and DFI behave as predicted i.e. there is a bias that grows quadratically with the decrease in the logarithm of the number of trials. Moreover, the QE correction was able to decrease the bias so that both DFI and rDFI are reliable with dataset of up to 8 times smaller than without the correction.

The last simulation that we conducted was based on the leaky integrate and fire neuronal model (Mazzoni et al., 2008) in order to generate data qualitatively similar to physiological recordings. We simulated two networks (Fig. 9B) virtually representing the information transfer scenario (Fig. 6A). As depicted in (Fig. 9A) DI shows a constant positive value representing transfer of information which is reasonable because the two networks were connected throughout the whole simulation. DFI shows a peak during the expected period of time when the stimulus signal was present in the simulation, however, it is negative throughout the whole simulation. Finally, rDFI forms a peak also during the period where it is expected and stays virtually zero otherwise.

4.2.2 Human Neurophysiological data

In order to confirm the behavior and show some of the benefits of the measure on real world data, we selected two datasets to which we applied it. We chose two dataset, both recorded during a task, composed of EEG and MEG recordings respectively. They are described in the Methods sections 4.1.3 and 4.1.4.

Information transfer during face detection (EEG)

Results from literature (Ince et al., 2016) have hypothesized that information about presence of an eye in the image is crucial for correct categorization and this information is transferred from the contra-lateral to the ipsi-lateral hemisphere with respect to the position of the eye (Fig. 15B). We tested such

hypothesis by means of rDFI analyses on an EEG dataset collected from human participants performing a face detection task (Rousselet et al., 2014). Subjects were asked to detect the presence of a face or a random texture while the image was covered by a bubble mask. Using rDFI, we were able to directly confirm the presence of the transfer and quantify it (Fig. 15A). We also confirmed, using a non-parametric statistical test (Maris & Oostenveld, 2007) that the transfer is the only significant transfer that is present in the data (Fig. 15D). Furthermore, we conducted comparisons between all 4 combinations of the transfer direction and the position of an eye, ipsi to contra-lateral and vice versa and left and right eye, respectively. Based on these tests, we confirmed that only transfers from contra-lateral to ipsy-lateral hemisphere are significant (Fig. 15C and F). Finally, we also tried to quantify the transfer using DI and DFI but they both, unlike rDFI, proved to be unable to capture the transfer (Fig. 15F).

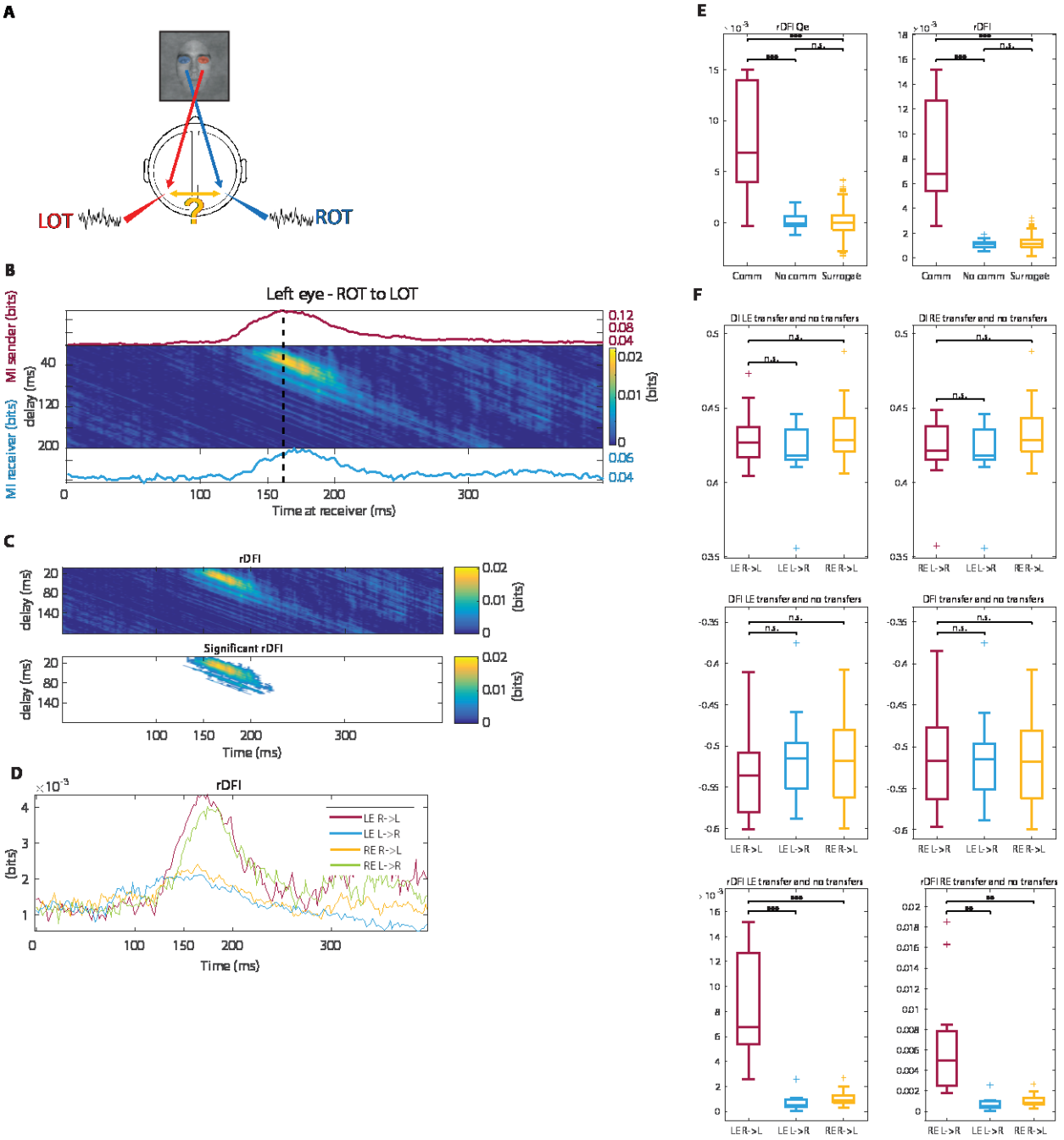

Figure 15 - The figure shows results of experiments based on the EEG dataset published by (Rousselet et al., 2014) that covers a task where subjects have to decide whether they see a face or a random texture in an image cover by a bubble mask. It was suggested that during the experiment, there is an information transfer about an eye (left LE or right RE) on the picture from the contralateral hemisphere to the ipsilateral hemisphere with respect to the eye ***A*** *(Ince et al., 2016).* ***B*** *captures the transfer of information about left eye from the right to the left hemisphere. The 2D map shows values for a given pair of time and delay that determines how far in the past the past values of* X *and* Y *were taken from. Line plots above and below show mutual information*

between the given occipito-temporal sensor (either sending or receiving one) and the visibility of the left eye. The panel clearly shows that rDFI identified a transfer at around 170ms as hypothesized by (Ince et al., 2016) that is in line with the peaks of mutual information. ***D*** *shows rDFI values on the same time line as in* ***B*** *averaged across all delays for all combinations of transfer direction and eye. There is a clear difference between the transfers from contra to ipsilateral, that form a visible peak around 170ms, whereas the opposite transfers do not.* ***C*** *In order to determine significance of rDFI values, we used a nonparametric statistical test introduced in (Maris & Oostenveld, 2007) that identifies all significant clusters of values in the 2D map. The upper plot shows the raw values of rDFI and the bottom one only the cluster that was identified as significant.* ***E*** *For the case of the left eye, right to left hemisphere transfer, we also compared the values in the cluster that captured the transfer (comm) to a cluster of the same shape and size but positioned elsewhere on the time line with no overlap (no comm) and surrogate values obtained by random shuffling of the stimulus values (left eye visibility). It shows that the values in the transfer cluster are significantly higher than those obtained from elsewhere and surrogates for both QE corrected and uncorrected rDFI.* ***F*** *shows comparison between the given contra to ipsilateral transfer (L – left, R - right) and the two ipsi to contralateral transfers for DI, DFI and rDFI and both eyes (LE and RE). It can be clearly observed that only rDFI was able to capture the transfer as DI and DFI does not show any difference between transfer directions for either eye.*

Information transfer in human visuomotor network (MEG)

In the second set of analyses, we used our novel metric to infer condition-dependent changes in information transfer over the human visuomotor network. Previous work has shown that the medial superior parietal cortex (SPCm) plays a key role in the visuomotor-related FC network (Brovelli et al., 2017). Briefly, two measures of centrality, such as the eigenvector centrality and the betweeness centrality, area flexibility, defined as the propensity of brain area to participate in multiple sub-networks, placed SPCm at the top of the cortical hierarchy. In particular, the SPCm was found to participate both in a visuo-parietal network early during the processing of visual information and during the planning of visuomotor associations with the dorsolateral premotor cortex (PMdl). Granger causality analyses also revealed a directional influence from the superior parietal lobe towards the premotor area (Brovelli et al., 2015).

We thus studied the DI, DFI and rDFI between the SPCm and the dorsolateral premotor cortex PMdl. (Fig. 16A - left panels) shows the increase in high-gamma activity (HGA) aligned on movement

onset. The DI and DFI showed a directional pattern that is consistent with previous results (higher directionality from the SPCm towards the PMdl), but the degree of asymmetry was weaker than the one observed with the rDFI. In fact, the rDFI was practically absent in the PMdl to SPCm direction and showed a clear drop after action execution (i.e., time equal to zero), as shown in (Fig. 16A - center and right panels). This result supports current knowledge about the dynamics of interaction between the superior parietal lobe and the dorsolateral premotor cortex, which is expected to occur over a timescale of tens to hundreds on milliseconds during a time interval prior to motor output.

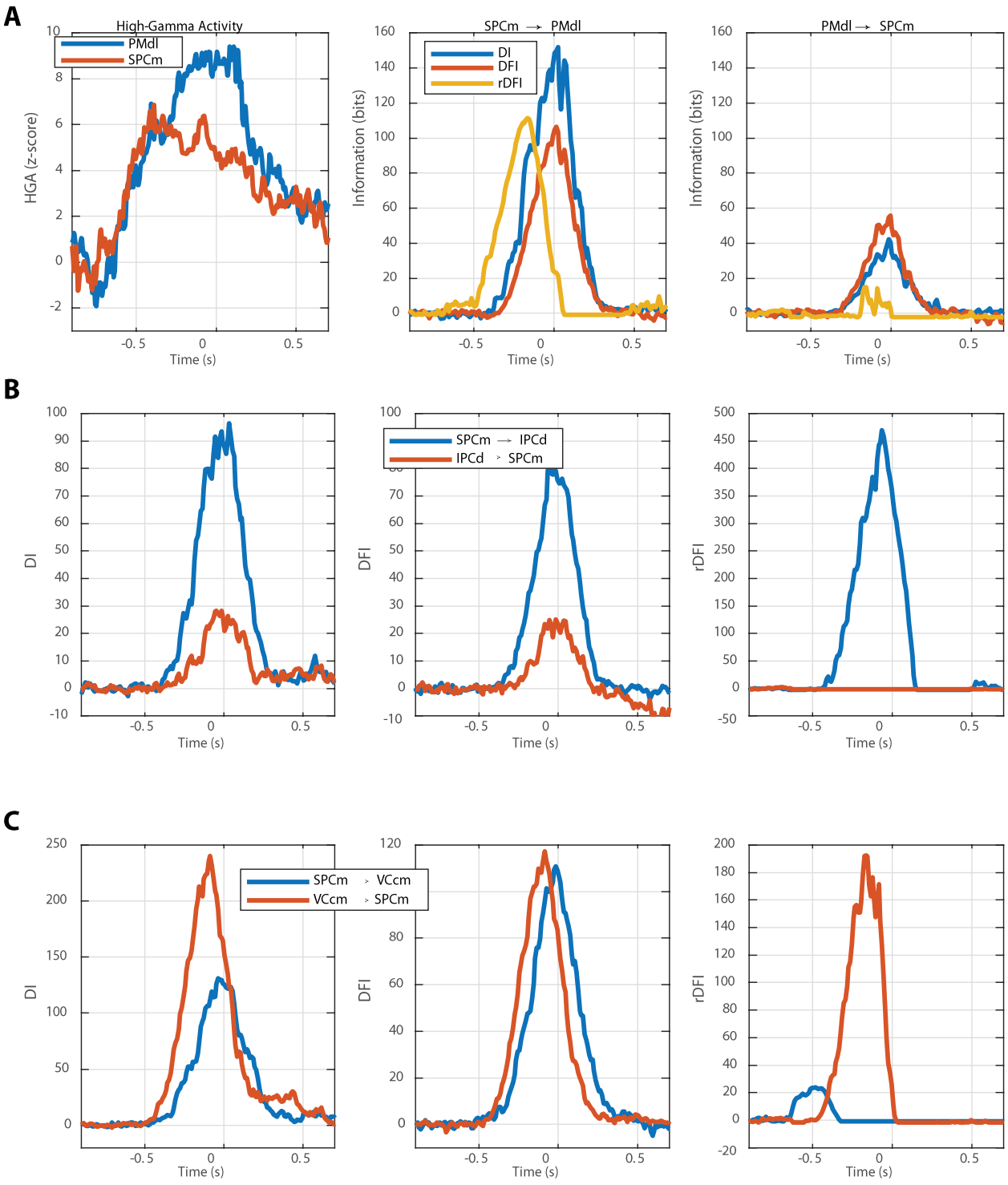

Figure 16 – (A) Left panel: time course of high-gamma activity (HGA) for the medial superior parietal cortex (SPCm) and the dorsolateral premotor cortex (PMdl), in red and blue, respectively. Central and right panels: directional influences (DI, DFI and rDFI) between the SPCm and PMdl. (B) Information flow between the SPCm and the dorsal Inferior Parietal Cortex (IPCd) for the DI (left panel), DFI (central panel) and rDFI

(right panel). (C) Information flow between the SPCm and the caudal medial Visual Cortex (VCcm) for the DI (left panel), DFI (central panel) and rDFI (right panel).

Then, we analyzed the coupling between the SPCm and a companion area in the parietal lobe, the dorsal Inferior Parietal Cortex (IPCd). This analysis was performed to test the efficiency of the rDFI in estimating efficiently directional coupling from neural signals from nearby brain regions that may share information and display common-driving effects. Contrary to what could be expected, rDFI displays a clear selectivity for one of the directionality of information flow, with respect to the DI and DFI (Fig. 16B).

Lastly, we analyzed the directionality of information flow between the SPCm and the caudal medial Visual Cortex (VCcm). The prediction was a directional information flow from the visual area to the parietal lobe during early stages of visuomotor integration. Indeed, as shown in Fig. 16C, rDFI showed a strong peak from the VCcm to the SPCm around 0.25s prior to button press. In addition, it is interesting to note that a reversal of information flow (from parietal to visual area) is observed around 0.5s prior to movement onset, thus corresponding roughly to the time interval of stimulus onset. Given the temporal structure of the task, we suggest that this increase of rDFI reflects top-down attentional processes that may be associated to the increased expectancy of the visual cue.

Overall, these results suggest that rDFI provides a significant measure condition-dependent information transfer able to extract features with both a temporal and directional selectivity, key elements for tracking cortico-cortical directional interactions.

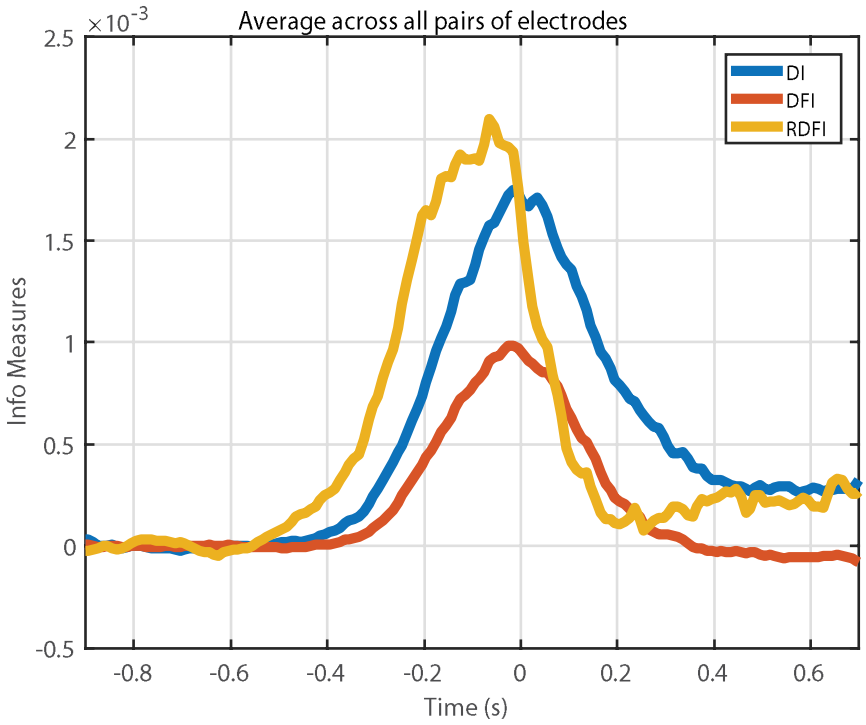

Figure 17 – Time course of information measures averaged across all pairs of brain areas (whole-brain analysis): DI (blue curve), DFI (red curve) and rDFI (yellow curve).

CHAPTER 5: DISCUSSION

A classical approach for inferring the amount of information transferred between neural signals exploits functional connectivity measures, such as Granger causality and Transfer Entropy (Bressler & Seth, 2011; Brovelli et al., 2004; Seth, Barrett, & Barnett, 2015; Wibral, Vicente, & Lizier, 2014). However, none of those measures carry information about the content of information. In our study, we developed a novel measure that quantifies information transfer about a specific feature based on the Partial Information Decomposition framework of (Williams & Beer, 2010).

The results of the simulations show that the rDFI manages to recover the underlying information transfer successfully. The first confirmation of its performance comes from the scenario of direct information transfer (Fig. 6) where it correctly captures the transfer and its behavior is virtually undistinguishable from that of DFI. Though unlike DFI, rDFI values remain correct even in other two scenarios in which DFI yielded significant negative values. These scenarios (Fig. 7 and 8) represent two nodes that do communicate, the sender node contains information about the feature of interest but does not communicate it. Additionally, we added an extra confounding signal (Fig. 8) providing the receiving node with information about a feature different than the one of interest but with the same probability distribution; however, rDFI still accurately represented that there was no transfer about the feature. Not only was rDFI robust to connectivity of the scenario but it also proved to be reliable even in presence of large amount of noise in the signal. In fact, even when we added so much noise into the signal that values of both DI and DFI turned insignificant, rDFI was still able to accurately quantify the transfer (Fig. 12). Ultimately, the desired behavior of rDFI is confirmed in the more complex simulation based on (Mazzoni et al., 2008)'s model which generates signal qualitatively much closer to that from real-world physiological

recording. In these simulations it was the only measure that exhibited a positive peak during the stimulus communication period and stayed negligibly small in the rest of the simulation.

rDFI demonstrated great sensitivity to the temporal evolution of information transfer. During simulations (Fig. 10 and 11) and the MEG experiment (Fig. 16 and 17) it exhibited asymmetrical information peaks that are unseen in DI or DFI. Based on the simulations we discovered that the abrupt decrease in rDFI values is linked to the delay determining from where the past values are obtained. If there is more information about the feature in past of the receiver than in past of the sender, rDFI rapidly goes to zero, correctly capturing the notion of the redundant information already being present in the receiver and thus not being part of the transfer. An ultimate case of this phenomenon is when the directionality of the transfer is assumed incorrectly, putting the information peak in the receiver earlier in time than that in the sender (Fig. 11). Due to its ability to correctly recognize the information already being present in receiver, rDFI remained zero in all scenarios with incorrect directionality, demonstrating very good sensitivity for it, unlike DFI and DI which showed a peak even in case of an "impossible" transfer.

The rDFI suffers from the sample size bias in the same way as all other Information Theory quantities and can be corrected by using Quadratic Extrapolation procedure (Panzeri et al., 2007; Strong et al., 1998) (Fig. 14). It reduces the amount of data that is necessary for precise values of rDFI to be obtained approximately 10 times. Therefore, we only need ≈8 times more data samples than the number of unique values of the full joint probability distribution that needs to be sampled. To remind the reader the number of those values is equal to the product of the numbers of possible values (numbers of bins) for stimulus, past of X, present of Y and past of Y. If there are less samples in the data, the values of rDFI might be overestimated.

Altogether, it suggests that rDFI not only provides information about the content of a transfer but that it can do so with either less data or data that are much noisier than those used with other transfer measures which is very important, especially in neuroscience where obtaining more data can be very difficult or impossible.

In order for the measure to be truly useful, solely its behavior does not suffice; there needs to be a statistical test that can be used with the method. We demonstrated, that significance of rDFI values can be tested by a nonparametric statistical test developed by (Maris & Oostenveld, 2007) (Fig. 15D) The test does not impose any assumptions on the data. This is very important trait of the test due to the transfer length and its delay are not a priori known. It can be used with two dimensions, when both time lags (in receiver and sender) are free parameters and one dimension when the values are aggregated across one of the lags.

Our method has few limitations. Probably the most restricting for practical use of the measure is its need of rather large dataset. The necessity comes from the need of reliable sampling of the full joint probability distribution of all three random variables of interest (past of sender, current values of receiver and past of receiver) and the stimulus as well. However, as demonstrated in the previous text it can be successfully mitigated by aggregating values of those random variables in a few buckets.

Another potential limitation resides in the assumption that I_{min} is a correct quantification of unique, redundant and synergistic information in a lattice approach. Given that our metric required three predictors (X_{t_2} , Y_{t_2} , Y_{t_1}), we could not use the measure proposed by (Bertschinger et al., 2014). It has been shown (Bertschinger et al., 2014; Griffith & Koch, 2014; Harder et al., 2013) that the potential incorrectness of I_{min} is due to miscategorizing purely unique information as synergistic, hence over estimating synergy

and potentially underestimating redundancy. Nevertheless, we have not observed this behavior in any of the simulations.

An alternative measure to quantify the novel information transfer from X to Y about S was proposed by Beer and Williams (Beer & Williams, 2015):

$$
\begin{aligned}
I_T(S; X_{t_1} \to Y_{t_2}) &= \mathrm{I}_{min}(S; \{Y_{t_2}\}\{X_{t_1}, Y_{t_1'}\}) - I_{min}(S; \{Y_{t_2}\}\{Y_{t_1'}\}) \\
&= \Pi_R(S; \{X_{t_1}\}\{Y_{t_2}\}) + \Pi_R(S; \{Y_{t_2}\}\{X_{t_1}, Y_{t_1'}\})
\end{aligned} \tag{5.1}
$$

The first Π_R term is our novel metric. The second term quantifies a different type of transfer that involves synergistic effects. It quantifies the information present in Y_{t_2} that is redundant with information not present in X_{t_1} alone, but that could be obtained synergistically combining X_{t_1} and Y_{t_1}, and furthermore is unique with respect to the information carried by Y_{t_1} alone.

Finally, it should be noted that the method is still a statistical measure (and this limitation holds for Granger causality and Transfer entropy as well) based solely on observation of events and therefore cannot guarantee causal relationship between two neuronal areas (neurons) (Panzeri, Harvey, Piasini, Latham, & Fellin, 2017). Consider scenario where two neurons seem to communicate information about the stimulus based on a statistical measure. It does not tell us whether there really was a communication, even indirectly, between the two neurons or whether they both received the same information from a third source with a different delay. In order to investigate such scenario and determine the precise causal relations between the neurons, one must intervene and be able to manipulate values of one of the neurons and then observe behavior of the rest of the system. However, the observation based methods are still very valuable and can uncover many dynamics of information exchange in the brain.

For what concerns the analysis of human neurophysiological data using the rDFI measure, we were able to verify presence of an information transfer from one to one to the other hemisphere, that was previously only hypothesized (Ince et al., 2016), about eye appearance in a picture based on an EEG data. To our knowledge, rDFI is the only measure capable of that at the moment.

The analysis of MEG data confirmed previous work combining single-trial high-gamma activity (HGA) from MEG recording and Granger causality analyses that showed that the performance of arbitrary visuomotor associations is characterized by an increase in functional connectivity (FC) over the sensorimotor and fronto-parietal network. The superior parietal area was found to play a driving role in the network, exerting Granger causality on the dorsal premotor area. Premotor areas acted as relay from parietal to medial prefrontal cortices, which played a receiving role in the network (Brovelli et al., 2015). More recently, a time-resolved analysis of FC revealed the presence of three partly-overlapping cortico-cortical and cortico-subcortical networks. First, visual and parietal regions coordinated with sensorimotor and premotor areas. Second, the dorsal fronto-parietal circuit together with the sensorimotor and associative fronto-striatal networks took the lead. Finally, cortico-cortical interhemispheric coordination among bilateral sensorimotor regions coupled with the left fronto-parietal network and visual areas (Brovelli et al., 2017). These results suggested that visuomotor mapping reflects the dynamic reconfiguration of multiple cortico-cortical and cortico-subcortical FC networks, displaying non-stationary, switching dynamics and areal flexibility over time scales relevant for task performance. However, previous work did quantify neither the informational content conveyed by large-scale directional influences nor their dynamics. To do so, we exploited our novel measure for the quantification of the directionality of information flow in a dynamical (time-resolved) manner between a key brain region in the superior parietal cortex.

As mentioned earlier, one of the key features of the rDFI is its ability to better detect the temporal evolution of information transfer with respect to the DI and DFI. Indeed, the results from the MEG analysis showed that the rDFI outperforms the DI and DFI in detecting the timing of information transfer between the superior parietal region and the dorsal premotor cortex. Indeed, the increase in rDFI from the SPCm to the PMdl (Fig. 16A, central panel) peaks approximately 0.2s before the motor response, therefore confirming the notion that the dorsal fronto-parietal network is crucial for the visuomotor computations transforming visual information into motor plans (Wise et al., 1996; Wise and Murray, 2000; Corbetta and Shulman, 2002; Culham and Valyear, 2006). The DI and DFI, however, are most probably influenced by the global increase in HGA observed around the motor response and their temporal dynamics decreases well after the motor output (Fig. 16B and C, the left and central panel). The difference in temporal evolution of the rDFI with respect to the DI and DFI is seen also at the large-scale level (Fig. 17).

The second main advantage of the rDFI is its ability to better detect directionality of information transfer with respect to the DI and DFI (Fig. 11). This was evident in the MEG results (Fig. 16B). For example, the bottom-up increase in information transfer expected to occur from the visual areas (VCcm) to the superior parietal cortex (SPCm) was evident in the rDFI, but was weaker or lacking for the DI and DFI, respectively (Fig. 16B). These results suggest that the rDFI reduces potential bias present in the DI and DFI due to the autocorrelation in the HGA. Whereas the DI and DFI show some bidirectional effect, it practically disappears for the rDFI.

Taken together, the higher selectivity for temporal and directional information results in the ability of the rDFI to discern rapid changes in directionality occurring over time, the so-called functional connectivity dynamics. This pattern of rapid change in directionality is shown in Fig. 16C (right panel).

Overall, the analysis of human neurophysiological data showed that the rDFI provides high discriminability of information transfer both in time and across directions. This suggests that rDFI has potential to uncover previously hidden dynamics in the information transfer between neuronal signals.

REFERENCES

Adey, W. R., Walter, D. O., & Hendrix, C. (1961). Computer techniques in correlation and spectral analyses of cerebral slow waves during discriminative behavior. *Experimental neurology, 3*(6), 501-524.

Aertsen, A., Gerstein, G., Habib, M., & Palm, G. (1989). Dynamics of neuronal firing correlation: modulation of" effective connectivity". *Journal of neurophysiology, 61*(5), 900-917.

Aertsen, A., & Preissl, H. (1991). Dynamics of activity and connectivity in physiological neuronal networks. *Nonlinear dynamics and neuronal networks, 2*, 281-301.

Aggleton, J., Keen, S., Warburton, E., & Bussey, T. (1997). Extensive cytotoxic lesions involving both the rhinal cortices and area TE impair recognition but spare spatial alternation in the rat. *Brain research bulletin, 43*(3), 279-287.

Alemi-Neissi, A., Rosselli, F. B., & Zoccolan, D. (2013). Multifeatural shape processing in rats engaged in invariant visual object recognition. *Journal of Neuroscience, 33*(14), 5939-5956.

Amblard, P.-O., & Michel, O. J. (2011). On directed information theory and Granger causality graphs. *Journal of computational neuroscience, 30*(1), 7-16.

Andermann, M. L., Kerlin, A. M., Roumis, D. K., Glickfeld, L. L., & Reid, R. C. (2011). Functional specialization of mouse higher visual cortical areas. *Neuron, 72*(6), 1025-1039.

Auzias, G., Coulon, O., & Brovelli, A. (2016). MarsAtlas: A cortical parcellation atlas for functional mapping. *Human brain mapping, 37*(4), 1573-1592.

Auzias, G., Lefevre, J., Le Troter, A., Fischer, C., Perrot, M., Régis, J., & Coulon, O. (2013). Model-driven harmonic parameterization of the cortical surface: HIP-HOP. *IEEE transactions on medical imaging, 32*(5), 873-887.

Barlow, J. S., & Brazier, M. A. (1954). A note on a correlator for electroencephalographic work. *Electroencephalography and clinical neurophysiology, 6*, 321-325.

Bastos, A. M., Usrey, W. M., Adams, R. A., Mangun, G. R., Fries, P., & Friston, K. J. (2012). Canonical microcircuits for predictive coding. *Neuron, 76*(4), 695-711.

Beer, R. D., & Williams, P. L. (2015). Information processing and dynamics in minimally cognitive agents. *Cognitive science, 39*(1), 1-38.

Bell, A. J. (2003). *The co-information lattice.* Paper presented at the Proceedings of the Fifth International Workshop on Independent Component Analysis and Blind Signal Separation: ICA.

Bertschinger, N., Rauh, J., Olbrich, E., Jost, J., & Ay, N. (2014). Quantifying unique information. *Entropy, 16*(4), 2161-2183.

Biswal, B., Zerrin Yetkin, F., Haughton, V. M., & Hyde, J. S. (1995). Functional connectivity in the motor cortex of resting human brain using echo-planar mri. *Magnetic resonance in medicine, 34*(4), 537-541.

Bosman, C. A., Schoffelen, J.-M., Brunet, N., Oostenveld, R., Bastos, A. M., Womelsdorf, T., . . . Fries, P. (2012). Attentional stimulus selection through selective synchronization between monkey visual areas. *Neuron, 75*(5), 875-888.

Bressler, S. L., & Menon, V. (2010). Large-scale brain networks in cognition: emerging methods and principles. *Trends in cognitive sciences, 14*(6), 277-290.

Bressler, S. L., & Seth, A. K. (2011). Wiener–Granger causality: a well established methodology. *Neuroimage, 58*(2), 323-329.

Brovelli, A., Badier, J.-M., Bonini, F., Bartolomei, F., Coulon, O., & Auzias, G. (2017). Dynamic reconfiguration of visuomotor-related functional connectivity networks. *Journal of Neuroscience, 37*(4), 839-853.

Brovelli, A., Chicharro, D., Badier, J.-M., Wang, H., & Jirsa, V. (2015). Characterization of cortical networks and corticocortical functional connectivity mediating arbitrary visuomotor mapping. *Journal of Neuroscience, 35*(37), 12643-12658.

Brovelli, A., Ding, M., Ledberg, A., Chen, Y., Nakamura, R., & Bressler, S. L. (2004). Beta oscillations in a large-scale sensorimotor cortical network: directional influences revealed by Granger causality. *Proceedings of the National Academy of Sciences of the United States of America, 101*(26), 9849-9854.

Buckner, R. L., Andrews-Hanna, J. R., & Schacter, D. L. (2008). The brain's default network: anatomy, function, and relevance to disease. *Ann N Y Acad Sci, 1124*, 1-38. doi:10.1196/annals.1440.011

Cadieu, C., Kouh, M., Pasupathy, A., Connor, C. E., Riesenhuber, M., & Poggio, T. (2007). A model of V4 shape selectivity and invariance. *Journal of neurophysiology, 98*(3), 1733-1750.

Callaway, E. M., & Marder, E. (2012). Common features of diverse circuits. *Curr Opin Neurobiol, 22*(4), 565-567. doi:10.1016/j.conb.2012.07.003

Coupé, P., Yger, P., Prima, S., Hellier, P., Kervrann, C., & Barillot, C. (2008). An optimized blockwise nonlocal means denoising filter for 3-D magnetic resonance images. *IEEE transactions on medical imaging, 27*(4), 425-441.

Cover, T. M., & Thomas, J. A. (2012). *Elements of information theory*: John Wiley & Sons.

DiCarlo, J. J., & Cox, D. D. (2007). Untangling invariant object recognition. *Trends in cognitive sciences, 11*(8), 333-341.

DiCarlo, J. J., Zoccolan, D., & Rust, N. C. (2012). How does the brain solve visual object recognition? *Neuron, 73*(3), 415-434.

Dosenbach, N. U., Fair, D. A., Miezin, F. M., Cohen, A. L., Wenger, K. K., Dosenbach, R. A., . . . Raichle, M. E. (2007). Distinct brain networks for adaptive and stable task control in humans. *Proceedings of the National Academy of Sciences, 104*(26), 11073-11078.

Douglas, R. J., Martin, K. A., & Whitteridge, D. (1989). A canonical microcircuit for neocortex. *Neural computation, 1*(4), 480-488.

Fox, M. D., & Raichle, M. E. (2007). Spontaneous fluctuations in brain activity observed with functional magnetic resonance imaging. *Nature reviews. Neuroscience, 8*(9), 700.

Fox, M. D., Snyder, A. Z., Vincent, J. L., Corbetta, M., Van Essen, D. C., & Raichle, M. E. (2005). The human brain is intrinsically organized into dynamic, anticorrelated functional networks. *Proceedings of the National Academy of Sciences of the United States of America, 102*(27), 9673-9678.

Friston, K. J., Harrison, L., & Penny, W. (2003). Dynamic causal modelling. *Neuroimage, 19*(4), 1273-1302.

Gallardo, L., Mottles, M., Vera, L., Carrasco, M. A., Torrealba, F., Montero, V. M., & Pinto-Hamuy, T. (1979). Failure by rats to learn a visual conditional discrimination after lateral peristriate cortical lesions. *Physiological Psychology, 7*(2), 173-177.

Garner, W. R. (1962). *Uncertainty and structure as psychological concepts.* Oxford, England: Wiley.

Gerstein, G. L., & Perkel, D. H. (1969). Simultaneously recorded trains of action potentials: analysis and functional interpretation. *Science, 164*(3881), 828-830.

Gevins, A. S., Doyle, J. C., Cutillo, B. A., Schaffer, R. E., Tannehill, R. S., & Bressler, S. L. (1985). Neurocognitive Pattern Analysis of a Visuospatial Task: Rapidly-Shifting Foci of Evoked Correlations Between Electrodes. *Psychophysiology, 22*(1), 32-43.

Geweke, J. F. (1984). Measures of conditional linear dependence and feedback between time series. *Journal of the American Statistical Association, 79*(388), 907-915.

Granger, C. W. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: Journal of the Econometric Society*, 424-438.

Greicius, M. D., Krasnow, B., Reiss, A. L., & Menon, V. (2003). Functional connectivity in the resting brain: a network analysis of the default mode hypothesis. *Proceedings of the National Academy of Sciences, 100*(1), 253-258.

Greicius, M. D., Supekar, K., Menon, V., & Dougherty, R. F. (2009). Resting-state functional connectivity reflects structural connectivity in the default mode network. *Cerebral cortex, 19*(1), 72-78.

Griffith, V., & Koch, C. (2014). Quantifying Synergistic Mutual Information. In M. Prokopenko (Ed.), *Guided Self-Organization: Inception* (pp. 159-190). Berlin, Heidelberg: Springer Berlin Heidelberg.

Hansen, P., Kringelbach, M., & Salmelin, R. (2010). *MEG: An introduction to methods*: Oxford university press.

Harder, M., Salge, C., & Polani, D. (2013). Bivariate measure of redundant information. *Physical Review E, 87*(1), 012130.

Hassabis, D., Kumaran, D., Summerfield, C., & Botvinick, M. (2017). Neuroscience-inspired artificial intelligence. *Neuron, 95*(2), 245-258.

Haxby, J. V., Connolly, A. C., & Guntupalli, J. S. (2014). Decoding neural representational spaces using multivariate pattern analysis. *Annual review of neuroscience, 37*, 435-456.

Hipp, J. F., Engel, A. K., & Siegel, M. (2011). Oscillatory synchronization in large-scale cortical networks predicts perception. *Neuron, 69*(2), 387-396.

Holmgren, C., Harkany, T., Svennenfors, B., & Zilberter, Y. (2003). Pyramidal cell communication within local networks in layer 2/3 of rat neocortex. *The Journal of physiology, 551*(1), 139-153.

Horwitz, B. (2003). The elusive concept of brain connectivity. *Neuroimage, 19*(2), 466-470.

Ince, R. A., Jaworska, K., Gross, J., Panzeri, S., Van Rijsbergen, N. J., Rousselet, G. A., & Schyns, P. G. (2016). The Deceptively Simple N170 Reflects Network Information Processing Mechanisms Involving Visual Feature Coding and Transfer Across Hemispheres. *Cerebral cortex, 26*(11), 4123-4135.

Ince, R. 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. *Scientific reports, 5*, 17681.

Kamiński, M., Ding, M., Truccolo, W. A., & Bressler, S. L. (2001). Evaluating causal relations in neural systems: Granger causality, directed transfer function and statistical assessment of significance. *Biological cybernetics, 85*(2), 145-157.

Klink, P. C., Dagnino, B., Gariel-Mathis, M.-A., & Roelfsema, P. R. (2017). Distinct Feedforward and Feedback Effects of Microstimulation in Visual Cortex Reveal Neural Mechanisms of Texture Segregation. *Neuron*.

Kriegeskorte, N., Goebel, R., & Bandettini, P. (2006). Information-based functional brain mapping. *Proceedings of the National Academy of Sciences of the United States of America, 103*(10), 3863-3868.

Lamme, V. A., & Roelfsema, P. R. (2000). The distinct modes of vision offered by feedforward and recurrent processing. *Trends in neurosciences, 23*(11), 571-579.

Liska, A., Galbusera, A., Schwarz, A. J., & Gozzi, A. (2015). Functional connectivity hubs of the mouse brain. *Neuroimage, 115*, 281-291.

Lizier, J. T., Heinzle, J., Horstmann, A., Haynes, J.-D., & Prokopenko, M. (2011). Multivariate information-theoretic measures reveal directed information structure and task relevant changes in fMRI connectivity. *Journal of computational neuroscience, 30*(1), 85-107.

Lu, H., Zou, Q., Gu, H., Raichle, M. E., Stein, E. A., & Yang, Y. (2012). Rat brains also have a default mode network. *Proceedings of the National Academy of Sciences, 109*(10), 3979-3984.

Mangin, J.-F., Riviere, D., Cachia, A., Duchesnay, E., Cointepas, Y., Papadopoulos-Orfanos, D., . . . Regis, J. (2004). A framework to study the cortical folding patterns. *Neuroimage, 23*, S129-S138.

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG-and MEG-data. *Journal of neuroscience methods, 164*(1), 177-190.

Massey, J. (1990). *Causality, feedback and directed information.* Paper presented at the Proc. Int. Symp. Inf. Theory Applic.(ISITA-90).

Mazzoni, A., Panzeri, S., Logothetis, N. K., & Brunel, N. (2008). Encoding of naturalistic stimuli by local field potential spectra in networks of excitatory and inhibitory neurons. *PLoS computational biology, 4*(12), e1000239.

McCormick, D. A., Connors, B. W., Lighthall, J. W., & Prince, D. A. (1985). Comparative electrophysiology of pyramidal and sparsely spiny stellate neurons of the neocortex. *Journal of neurophysiology, 54*(4), 782-806.

McDaniel, W. F., Coleman, J., & Lindsay, J. F. (1982). A comparison of lateral peristriate and striate neocortical ablations in the rat. *Behavioural brain research, 6*(3), 249-272.

McGill, W. (1954). Multivariate information transmission. *Transactions of the IRE Professional Group on Information Theory, 4*(4), 93-111.

Mitra, P. P., & Pesaran, B. (1999). Analysis of dynamic brain imaging data. *Biophysical journal, 76*(2), 691-708.

Nalatore, H., Ding, M., & Rangarajan, G. (2007). Mitigating the effects of measurement noise on Granger causality. *Physical Review E, 75*(3), 031123.

Nolte, G. (2003). The magnetic lead field theorem in the quasi-static approximation and its use for magnetoencephalography forward calculation in realistic volume conductors. *Physics in medicine and biology, 48*(22), 3637.

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.

Panzeri, S., Macke, J. H., Gross, J., & Kayser, C. (2015). Neural population coding: combining insights from microscopic and mass signals. *Trends in cognitive sciences, 19*(3), 162-172.

Panzeri, S., Schultz, S. R., Treves, A., & Rolls, E. T. (1999). Correlations and the encoding of information in the nervous system. *Proceedings of the Royal Society of London B: Biological Sciences, 266*(1423), 1001-1012.

Panzeri, S., Senatore, R., Montemurro, M. A., & Petersen, R. S. (2007). Correcting for the sampling bias problem in spike train information measures. *Journal of neurophysiology, 98*(3), 1064-1072.

Panzeri, S., & Treves, A. (1996). Analytical estimates of limited sampling biases in different information measures. *Network: Computation in Neural Systems, 7*, 87-107.

Percival, D. B., & Walden, A. T. (1993). *Spectral analysis for physical applications*: Cambridge University Press.

Pereda, E., Quiroga, R. Q., & Bhattacharya, J. (2005). Nonlinear multivariate analysis of neurophysiological signals. *Progress in neurobiology, 77*(1), 1-37.

Perrot, M., Rivière, D., & Mangin, J.-F. (2011). Cortical sulci recognition and spatial normalization. *Medical image analysis, 15*(4), 529-550.

Pola, G., Thiele, A., Hoffmann, K., & Panzeri, S. (2003). An exact method to quantify the information transmitted by different mechanisms of correlational coding. *Network-Computation in Neural Systems, 14*(1), 35-60.

Raichle, M. E. (2015). The restless brain: how intrinsic activity organizes brain function. *Phil. Trans. R. Soc. B, 370*(1668), 20140172.

Riesenhuber, M., & Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nat Neurosci, 2*(11), 1019-1025.

Roebroeck, A., Formisano, E., & Goebel, R. (2005). Mapping directed influence over the brain using Granger causality and fMRI. *Neuroimage, 25*(1), 230-242.

Rosselli, F. B., Alemi, A., Ansuini, A., & Zoccolan, D. (2015). Object similarity affects the perceptual strategy underlying invariant visual object recognition in rats. *Frontiers in neural circuits, 9*.

Rousselet, G. A., Ince, R. A., van Rijsbergen, N. J., & Schyns, P. G. (2014). Eye coding mechanisms in early human face event-related potentials. *Journal of vision, 14*(13), 7-7.

Sato, J. R., Junior, E. A., Takahashi, D. Y., de Maria Felix, M., Brammer, M. J., & Morettin, P. A. (2006). A method to produce evolving functional connectivity maps during the course of an fMRI experiment using wavelet-based time-varying Granger causality. *Neuroimage, 31*(1), 187-196.

Schreiber, T. (2000). Measuring Information Transfer. *Physical Review Letters, 85*(2), 461-464.

Serre, T., Oliva, A., & Poggio, T. (2007). A feedforward architecture accounts for rapid categorization. *Proceedings of the National Academy of Sciences, 104*(15), 6424-6429.

Serre, T., Wolf, L., Bileschi, S., Riesenhuber, M., & Poggio, T. (2007). Robust object recognition with cortex-like mechanisms. *IEEE transactions on pattern analysis and machine intelligence, 29*(3), 411-426.

Seth, A. K., Barrett, A. B., & Barnett, L. (2015). Granger causality analysis in neuroscience and neuroimaging. *Journal of Neuroscience, 35*(8), 3293-3297.

Shannon, C. (1948). A mathematical theory of communication. *The Bell System Technical Journal, 27*(3), 379-423.

Sjöström, P. J., Turrigiano, G. G., & Nelson, S. B. (2001). Rate, timing, and cooperativity jointly determine cortical synaptic plasticity. *Neuron, 32*(6), 1149-1164.

Strong, S. P., Koberle, R., van Steveninck, R. R. d. R., & Bialek, W. (1998). Entropy and information in neural spike trains. *Physical Review Letters, 80*(1), 197.

Studený, M., & Vejnarová, J. (1998). The multiinformation function as a tool for measuring stochastic dependence. *Learning in graphical models,* pp. 261-297.

Tafazoli, S., Safaai, H., De Franceschi, G., Rosselli, F. B., Vanzella, W., Riggi, M., . . . Zoccolan, D. (2017). Emergence of transformation-tolerant representations of visual objects in rat lateral extrastriate cortex. *eLife, 6*, e22794.

Tononi, G., Sporns, O., & Edelman, G. M. (1994). A measure for brain complexity: relating functional segregation and integration in the nervous system. *Proceedings of the National Academy of Sciences, 91*(11), 5033-5037.

Tou, J. T., & González, R. C. (1974). *Pattern recognition principles*: Addison-Wesley Pub. Co.

Tuckwell, H. (1988). *Introduction to Theoretical Neurobiology: Volume 1, Linear Cable Theory and Dendritic Structure (Cambridge Studies in Mathematical Biology)*: Cambridge University Press.

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. *Proceedings of the National Academy of Sciences, 111*(40), 14332-14341.

Van Kerkoerle, T., Self, M. W., & Roelfsema, P. R. (2017). Layer-specificity in the effects of attention and working memory on activity in primary visual cortex. *Nature communications, 8*.

Van Veen, B. D., Van Drongelen, W., Yuchtman, M., & Suzuki, A. (1997). Localization of brain electrical activity via linearly constrained minimum variance spatial filtering. *IEEE Transactions on biomedical engineering, 44*(9), 867-880.

Varela, F., Lachaux, J.-P., Rodriguez, E., & Martinerie, J. (2001). The brainweb: phase synchronization and large-scale integration. *Nature reviews neuroscience, 2*(4), 229-239.

Vicente, R., Wibral, M., Lindner, M., & Pipa, G. (2011). Transfer entropy—a model-free measure of effective connectivity for the neurosciences. *Journal of computational neuroscience, 30*(1), 45-67.

Vinken, K., Vermaercke, B., & de Beeck, H. P. O. (2014). Visual categorization of natural movies by rats. *Journal of Neuroscience, 34*(32), 10645-10658.

von der Malsburg, C., Phillips, W. A., & Singer, W. (2010). *Dynamic coordination in the brain: from neurons to mind*: MIT Press.

Watanabe, S. (1960). Information theoretical analysis of multivariate correlation. *IBM Journal of research and development, 4*(1), 66-82.

Wibral, M., Vicente, R., & Lizier, J. T. (2014). *Directed information measures in neuroscience*: Springer.

Wiener, N. (1956). The theory of prediction. In E. F. Beckenbach (Ed.), *Modern mathematics for engineers* (Vol. 1). New York: McGraw-Hill.

Williams, P. L., & Beer, R. D. (2010). Nonnegative decomposition of multivariate information. *arXiv preprint arXiv:1004.2515*.

Zoccolan, D., Oertelt, N., DiCarlo, J. J., & Cox, D. D. (2009). A rodent model for the study of invariant visual object recognition. *Proceedings of the National Academy of Sciences, 106*(21), 8748-8753.