Identification of causal relations in neuroimaging data with latent confounders: An instrumental variable approach

Moritz Grosse-Wentrup[1], Dominik Janzing[1], Markus Siegel[2,3] and Bernhard Schölkopf[1]

[1] *Department Empirical Inference, Max Planck Institute for Intelligent Systems, Spemannstr. 38, 72076 Tübingen, Germany;* [2,3] *Centre for Integrative Neuroscience and MEG Center, Eberhard-Karls Universität Tübingen, Otfried-Müller-Str. 47, 72076 Tübingen, Germany*

Abstract

We consider the task of inferring causal relations in brain imaging data with latent confounders. Using a priori knowledge that randomized experimental conditions can not be effects of brain activity, we derive statistical conditions that are sufficient for establishing a causal relation between two neural processes, even in the presence of latent confounders. We provide an algorithm to test these conditions on empirical data, and illustrate its performance on simulated as well as on experimentally recorded EEG data.

1. Introduction

Inferring the causal structure of a cortical network is a central goal in neuroimaging [1]. Various methods have been developed to infer causal relations from brain imaging data, including structural equation modeling (SEM) [2, 3], Granger causality (GC) [4, 5, 6], dynamic causal modeling (DCM) [7, 8], and

Corresponding author: M. G.-W. (email: moritzgw@tuebingen.mpg.de; phone: +49-7071-601-547). M. S. was supported by the Centre for Integrative Neuroscience (Deutsche Forschungsgemeinschaft, EXC 307). The authors would like to thank Sebastian Weichwald for comments on an earlier version of this article.

causal Bayesian networks (CBNs) [9, 10, 11, 12, 13]. These methods commonly assume *causal sufficiency*; that is, they presume that all causally relevant variables have been observed. This assumption is often implausible, because various factors can confound a causal analysis. These factors include, but are not limited to, unmeasured brain regions in an fMRI analysis [2, 8, 14], cardio-ballistic artifacts in ECoG recordings [15], and volume conduction of cortical and non-cortical current sources in EEG or MEG data [16, 17]. Because it is not trivial to anticipate potential confounders, results obtained with methods based on causal sufficiency must be interpreted with caution.

Latent confounders can be addressed by the IC* [18] and FCI algorithms [19, 20], which use the theory of ancestral graphs. Theoretically, both algorithms can distinguish genuine causal relations from spurious relations induced by latent confounders. In practice, the involved statistical tests are complex, which currently limits their application in neuroimaging to variables that are jointly Gaussian distributed [21]. The assumption of jointly Gaussian distributed variables has been criticized as unreasonable for neuroimaging data [22, 23, 12].

We contribute to research on causal inference with latent confounders in two ways. First, we show that the statistical tests required to identify a genuine causal relation can be simplified when the experimental condition is randomized. Using the a priori knowledge that a randomized experimental condition cannot be caused by neural processes, we analytically prove that if two neural processes are modulated by an experimental condition, a single test of conditional independence is sufficient to establish a genuine causal relation between those processes. To emphasize the requirement that, in our approach, the experimental conditions must be randomized, we later refer to them as the stimuli presented to a subject. Second, by using linear regression, we reduce the required conditional independence test to a marginal independence test. This test is advantageous because asymptotically consistent statistical tests are readily available for marginal independence [24, 25, 26], but not for conditional independence [27, 28]. We prove that this linearized conditional independence test is sufficient but not necessary for conditional independence: while our test may

fail to detect conditional independence if the assumption of linearity is not met, a positive test result implies that this assumption has been fulfilled. Taken together, our two contributions lead to a non-parametric version of the instrumental variable approach to causal inference [29, 18]. The resulting algorithm, which we term *stimulus-based causal inference* (SCI), can provide empirical evidence for a causal relation between two neural processes, even in the presence of latent confounders.

We demonstrate the performance of the SCI algorithm on simulated as well as on experimentally recorded EEG data. We first use a neural mass model for spectral responses in electrophysiology [30] to provide estimates of the power and of the false discovery rate (FDR) of the SCI algorithm for a variety of causal models. We then show how our method can be used to infer group-level causal relations on EEG data, which we recorded for a study on brain-computer interfacing (BCI) [31]. In this study, subjects were trained via neurofeedback to self-regulate the amplitude of γ-oscillations (55–85 Hz) in the right superior parietal cortex (SPC), a primary node of the central executive network (CEN) [32]. Because transcranial magnetic stimulation (TMS) of the CEN has been found to modulate the medial prefrontal cortex (MPC) [33], we hypothesized that self-regulation of γ-power in the right SPC causes variations in γ-power in the MPC. Consistent with this hypothesis, the SCI algorithm determined the MPC to be modulated by the right SPC. We conclude the article with a discussion of the utility and of the limitations of causal inference to study the structure and the function of cortical networks.

We note that the SCI algorithm is applicable not only to EEG recordings but also to any neuroimaging data set that is based on randomized experimental conditions. We have condensed the SCI algorithm into one line of Matlab code, which is available at `http://brain-computer-interfaces.net`.

2. Methods

We begin this section by introducing the framework of causal Bayesian networks (CBNs), which our work is based on (cf. [9, 10, 11, 12, 13] for applications of this framework in neuroimaging). We then present the sufficient conditions to establish causal influence of one cortical process on another in stimulus-based experiments (subsection 2.2). In subsection 2.3, we use linear regression to reduce the required conditional independence test to a marginal independence test. We discuss how to apply the resulting causal inference procedure to empirical data in subsection 2.4. We conclude the methods section with a discussion of the relation of the SCI algorithm to instrumental variables in subsection 2.5.

2.1. Causal Bayesian networks

In the framework of CBNs, a random variable x is a cause of another random variable y if setting x to different values by an external intervention changes the probability distribution over y [18, 19]. In the notation of the *do*-calculus, this is expressed as $p(y|\mathrm{do}(x)) \neq p(y)$ for some values of x and y. Thus, the framework of CBNs defines cause-effect relations in terms of the impact of external manipulations. This definition contrasts those of frameworks which define causality in terms of information transfer [4, 34, 6, 35].

Causal relations between a set $\mathcal{X}$ of random variables are represented by edges in a directed acyclic graph (DAG). The *causal Markov condition* (CMC) relates the structure of a DAG, as represented by its edges, to statistical independence relations between the variables in $\mathcal{X}$. Specifically, it states that every (conditional) independence implied by a DAG is also found in the joint probability distribution $p(\boldsymbol{x})$. We recall that two random variables x and y are statistically independent (conditional on a third random variable z) if and only if their joint distribution factorizes into the product of its marginals, i.e. if and only if $p(x, y) = p(x)p(y)$ ($p(x, y|z) = p(x|z)p(y|z)$). Intuitively, this states that observing x does not provide any information on how likely certain outcomes of y are (and vice versa). We abbreviate statistical independence between x and

y (conditional on z) as $x \perp\!\!\!\perp y$ ($x \perp\!\!\!\perp y|z$). Assuming the CMC, (conditional) independence relations can be read off the structure of a DAG by checking for *d-separation* properties. A set of nodes $\mathcal{D}$ is said to d-separate x and y if every path from x to y contains at least one variable z such that either z is a collider ($\rightarrow z \leftarrow$) and no descendant of z (including z itself) is in $\mathcal{D}$; or z is not a collider and z is in $\mathcal{D}$. We provide examples of d-separation in the next paragraph and refer the interested reader to [18] or [19] for a more exhaustive introduction to the concept of d-separation. The CMC thus relates structural properties of DAGs to empirically observable independence relations. To perform causal inference, we also need to relate empirically observable independence relations to structural properties of the data-generating DAG. This is achieved by the assumption of *faithfulness*. Faithfulness asserts that every (conditional) independence relation in $p(\boldsymbol{x})$ is implied by the structure of the associated DAG. Taken together, the CMC and faithfulness ensure that two variables x and y are conditionally independent given z if and only if x and y are d-separated by z. This equivalence gives us insight into the structure of a DAG from empirically testable (conditional) independence relations.

We now provide three examples of d-separation that are relevant to our following arguments. First, consider the chain $x \rightarrow z \rightarrow y$. Here, x and y are marginally dependent ($x \not\perp\!\!\!\perp y$), because x influences y via z. However, as z d-separates x and y by blocking the directed path from x to y, x and y are statistically independent given z ($x \perp\!\!\!\perp y|z$). Second, consider the fork $x \leftarrow z \rightarrow y$. Again, x and y are marginally dependent ($x \not\perp\!\!\!\perp y$), because they share a common cause z. This common cause z again d-separates x and y by removing the joint effect of z on x and y, rendering x and y independent conditional on z ($x \perp\!\!\!\perp y|z$). Third, consider the collider $x \rightarrow z \leftarrow y$. In this case, x and y are independent ($x \perp\!\!\!\perp y$), because they are d-separated by the empty set. Because z is a joint effect of x and y, however, it unblocks the previously blocked path between x and y, rendering x and y dependent conditional on z ($x \not\perp\!\!\!\perp y|z$).

These three examples form the basis of causal inference in CBNs. For in-

stance, if we observe that $x \perp\!\!\!\perp y$ yet $x \not\perp\!\!\!\perp y|z$, then we can conclude that our data has not been generated by a chain or by a fork. These observations limit the possible causal structures to only collider and DAGs with additional (latent) variables. A more comprehensive introduction to the framework of CBNs in the context of neuroimaging is given in [12].

2.2. Causal inference in stimulus-based paradigms

In this article, we only consider DAGs over a set of three random variables, $\mathcal{V} = \{s, x, y\}$. The variables x and y represent brain state features, and s represents an experimental condition. For our theoretical arguments, we assume the joint probability distribution $p(s, x, y)$ to be known. This assumption implies that we have access to an oracle for any conditional independence relation in $\mathcal{V}$. We relax this assumption in section 2.4. Note that, while x and y may represent any measure of brain activity, it is helpful to consider trial-averaged blood-oxygen-level-dependent (BOLD) activity at different cortical locations or trial-averaged bandpower at two EEG channels as examples.

In the following, we assume that s codes a randomized experimental stimulus that is presented to the subject before x and y are measured. This assumption leads to the following theorem.

Theorem 1 (Causal inference in stimulus-based paradigms)**.** *Let s, x, and y be three random variables with a joint probability distribution $p(s, x, y)$ that is faithful to its generating DAG. Further, assume that s codes a randomized experimental stimulus that is presented before x and y are measured. Then the following three conditions are sufficient for x to be a genuine cause of y ($x \to y$):*

1. *s is not independent of x ($s \not\perp\!\!\!\perp x$),*
2. *s is not independent of y ($s \not\perp\!\!\!\perp y$), and*
3. *s and y are independent conditional on x ($s \perp\!\!\!\perp y|x$).*

Proof. To illustrate why theorem 1 is sufficient but not necessary for x to be a genuine cause of y, we first assume causal sufficiency; that is, we rule out

the presence of latent confounders. We then extend the proof to also consider unmeasured common causes.

Because s is a randomized experimental stimulus that is presented before measuring x, any dependence between s and x ($s \not\perp\!\!\!\perp x$) implies that s is a (not necessarily direct) cause of x [36]. This argument also applies to the relation between s and y. Conditions one and two thus ensure that $p(s, x, y)$ is generated by a DAG in which s is a cause of x and of y. Figure 1 shows all possible DAGs for the variable set $\{s, x, y\}$ for which this is the case. In DAG A, the influence of s on y is mediated by x. As such, x d-separates s and y, which implies $s \perp\!\!\!\perp y|x$. This establishes that the DAG $s \to x \to y$, in which x is a cause of y, is consistent with the three conditions of theorem 1. Next, we must show that DAG A is the only DAG consistent with these three conditions. This is easily seen by noting that DAGs B–E contain an arrow from s to y. As such, in these DAGs x does not d-separate s and y, so conditioning on x does not render s and y independent. This completes the proof of theorem 1 under the assumption of causal sufficiency.

To extend the proof to also allow for latent variables, note that under the assumption of faithfulness the condition $s \perp\!\!\!\perp y|x$ implies that in the true DAG – i.e., the (potentially latent) structure that generated $p(s, x, y)$ – x d-separates s and y. This implication means every directed path from s to y must be intersected by x. Because s is a cause of y, at least one such path exists and contains a directed subpath from x to y (i.e., $x \to y$). □

We emphasize that the conditions in theorem 1 are only sufficient but not necessary for x to be a genuine cause of y. Indeed, DAG D in figure 1 is an example in which x is a cause of y yet $s \perp\!\!\!\perp y|x$ does not hold. We further note that the theorem actually only requires the following implication of faithfulness: Given three variables s, x, y with $s \perp\!\!\!\perp y|x$, then every directed path from s to y contains x as an intermediate node. This condition is close in spirit to partial faithfulness defined for linear models in [37].

We further note that s may represent stimuli of various complexity. For

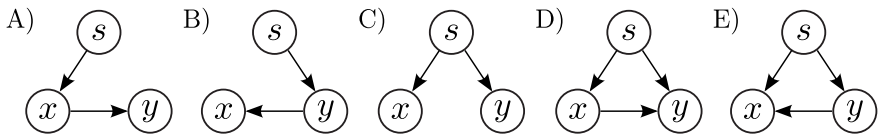

Figure 1: All DAGs for the set of variables $\{s, x, y\}$ in which s is a cause of both x and y. Note that we do not need to consider DAGs in which s is not a cause of both x and y, because these DAGs are ruled out by conditions one and two of theorem 1

instance, s may encode whether a house or a face is shown to a subject. In this case, x could be chosen to represent brain activity in a part of the visual cortex that differentially responds to houses vs. faces. We could then use theorem 1 to investigate if brain activity in a higher cortical area, represented by y, is modulated by x. It is also admissible, however, for s to represent more complex stimuli. For instance, s may represent the visual instruction to carry out one of two cognitive tasks, such as performing a mathematical computation vs. recalling a positive memory. While it may be less straightforward to determine the functional roles of x and y in this setting, e.g. whether x represents the visual instruction or the actual execution of the cognitive task, theorem 1 can provide insights into the causal relation between the neural states as long as s is randomized and presented to the subject before x and y are measured.

2.3. Regression-based conditional independence tests

While the first two conditions of theorem 1 are straightforward to test on empirical data, the third condition requires a non-trivial conditional independence test. Conditional independence tests are complex because it is difficult to sample from the corresponding null-distribution [28]. We circumvent this problem by using linear regression to convert a conditional to a marginal independence test. This approach is based on the following theorem.

Theorem 2 (Regression-based conditional independence tests)**.** *Let s, x, and y be three random variables with joint probability distribution $p(s, x, y)$. If there exists a function $f(x)$ s.t. $y - f(x) \perp\!\!\!\perp (s, x)$, then $s \perp\!\!\!\perp y|x$.*

Proof. The mutual information of s and y given x can be expressed as

$$I(s; y|x) = I(s; y - f(x)|x) = I(y - f(x); (s, x)) - I(y - f(x); x). \qquad (1)$$

Here, the first equality follows from the property that adding constants does not affect mutual information, and the second equality follows from the chain rule for mutual information. If $y - f(x) \perp\!\!\!\perp (s, x)$, then both terms on the right side of (1) are zero. This implies that $I(s; y|x) = 0$ and hence that $s \perp\!\!\!\perp y|x$. □

Instead of testing $s \perp\!\!\!\perp y|x$, we only need to check whether there exists a function $f(x)$ for which $y - f(x) \perp\!\!\!\perp (s, x)$. This method is advantageous for two reasons. First, asymptotically consistent conditional independence tests are not readily available, but marginal independence tests are [24, 25, 26]. Second, it is sufficient to find one function for which $y - f(x) \perp\!\!\!\perp (s, x)$ to conclude that $s \perp\!\!\!\perp y|x$. Specifically, we show in section 3 that it is often sufficient to consider linear functions of the form $f(x) = ax + b$. In contrast to causal inference methods based on partial correlation [21], our method does not require the questionable assumption of joint Gaussianity. Naturally, this advantage comes at a cost: Because the conditions of the regression-based conditional independence test are only sufficient but not necessary for $s \perp\!\!\!\perp y|x$, our test may fail to find certain types of conditional independence.

2.4. The stimulus-based causal inference (SCI) algorithm

In this section, we discuss how to apply theorems 1 and 2 to empirical data. If we had access to an oracle for conditional independence relations, it would be straightforward to test the three conditions of theorem 1. In practice, however, we only have access to a set $\{\mathcal{V}_1, \ldots, \mathcal{V}_N\}$ of N samples $\mathcal{V}_n = \{s_n, x_n, y_n\}$, which we assume to be independent and identically distributed (i.i.d.). Testing the conditions of theorem 1 on this data set requires three statistical tests.

First, we must test whether $s \perp\!\!\!\perp x$; that is, we need to test whether the brain state feature x is modulated by the experimental stimuli. This univariate statistical test for independence can be done by using any established method for a given data modality; for example, a general linear model can be used

for fMRI data [38], and a t-test can be used for bandpower changes in EEG data [39]. Here, we choose an ordinary correlation analysis. In this analysis, we first compute Pearson's correlation coefficient $\rho_{s,x}$ between s and x, and then estimate the p-value under the null-hypothesis $H0_1 : s \perp\!\!\!\perp x$ by randomly permuting the trial order of x 10^4 times. We count the instances in which the absolute value of the resulting correlation coefficient $\rho_{s,x}^{H0}$ exceeds the absolute value of $\rho_{s,x}$, and reject $H0_1$ if we estimate the probability of $|\rho_{s,x}^{H0}| > |\rho_{s,x}|$ to be less than $\alpha_{\text{rej}} = 0.01$, i.e. if $p < \alpha_{\text{rej}}$. Second, we test the second null-hypothesis $H0_2 : s \perp\!\!\!\perp y$ using the same test procedure as for $H0_1$.

Third, we test whether $s \perp\!\!\!\perp y|x$. Based on theorem 2, we first perform a linear regression from x to y. We determine the slope a and intercept b that minimize the sample variance of the residuals $\hat{y}_n = y_n - ax_n - b$. We then use the Hilbert–Schmidt independence criterion (HSIC) to test the null-hypothesis $H0_3 : \hat{y} \perp\!\!\!\perp (s, x)$ [25]. For this step, it is essential to use a non-linear test for statistical independence, rather than a correlation analysis, because zero correlation only implies statistical independence under the restrictive assumption of jointly Gaussian distributed variables. We apply the HSIC by first computing $\text{HSIC}_{\hat{y},(s,x)}$ between $\hat{y}$ and (s, x), using a Gaussian kernel with the kernel width set to the median distance between points in input space. We then estimate the p-value under the null-hypothesis by randomly permuting the trial order of $\hat{y}$ 10^4 times. We accept $H0_3$ if we estimate the probability of $\text{HSIC}_{\hat{y},(s,x)}^{H0} > \text{HSIC}_{\hat{y},(s,x)}$ to be greater than $\alpha_{\text{acc}} = 0.25$; i.e. if $p > \alpha_{\text{acc}}$. The parameters α_{rej} and α_{acc} jointly control the power and the false discovery rate (FDR) of the SCI algorithm. We justify our choice of these parameters in section 3.1. If we reject $H0_1$ and $H0_2$ and accept $H0_3$, we consider the conditions of theorem 1 fulfilled and conclude that x is a genuine cause of y. A summary of the SCI algorithm is given in table 1. Matlab code of the SCI algorithm is available at `http://brain-computer-interfaces.net`.

We note that any non-linear test for statistical independence can be used to test $H0_3 : \hat{y} \perp\!\!\!\perp (s, x)$. We have chosen the HSIC with a Gaussian kernel and the kernel width set to the median distance between points in input space for three

Input: N i.i.d. samples $\{s_n, x_n, y_n\}$ of two brain state features x and y and one randomized experimental stimulus s, presented prior to recording of x and y.

Procedure:

1. Choose significance level α_{rej} for rejecting independence; e.g. $\alpha_{\text{rej}} = 0.01$.
2. Choose significance level α_{acc} for accepting independence; e.g. $\alpha_{\text{acc}} = 0.25$.
3. Test $H0_1 : s \perp\!\!\!\perp x$ by using a correlation analysis. Reject $H0_1$ if the resulting p-value is smaller than α_{rej}.
4. Test $H0_2 : s \perp\!\!\!\perp y$ by using a correlation analysis. Reject $H0_2$ if the resulting p-value is smaller than α_{rej}.
5. Perform a linear regression from x to y and compute the residuals $\hat{y}$.
6. Test $H0_3 : \hat{y} \perp\!\!\!\perp (s, x)$ using a non-linear independence test such as the HSIC [25]. Accept $H0_3$ if the resulting p-value is larger than α_{acc}.

Result: If $H0_1$ and $H0_2$ are rejected and $H0_3$ is accepted, conclude that x is a genuine cause of y: $x \to y$.

Table 1: Summary of the SCI algorithm.

reasons. First, the Gaussian kernel is a universal kernel [40]. This guarantees that it detects any dependence in the large-sample setting [26]. Second, the Gaussian kernel has been empirically found to work well in a wide variety of settings [41]. And third, we chose the heuristic proposed in [25] for the kernel width to keep the number of statistical tests in section 3 computationally tractable.

We further note that it is not possible to mathematically quantify the probability that the SCI algorithm returns a false positive test result, e.g. in analogy to a p-value in a traditional statistical analysis, because the SCI algorithm requires the acceptance of a null-hypothesis to conclude that x is a cause of y. We address this problem in section 3.1 by providing empirical estimates of the FDR of the SCI algorithm in a variety of experimental settings.

2.5. Relation to instrumental variables

Instrumental variables (IVs) are used to estimate the causal effect of x on y in settings in which interventions on x are not possible, e.g. due to ethical considerations or technical constraints [29]. They have a long history in econometrics [42, 43] and have more recently been applied in the context of genetics [44, 45]. A variable z qualifies as an IV for the pair $\{x, y\}$ if the following three conditions are fulfilled: a) z is statistically independent of all joint (and potentially hidden) common causes of x and y; b) z is not independent of x; and c) the effect of z on y is mediated solely by x. The SCI algorithm can be interpreted as a statistical test whether s qualifies as an IV for the pair $\{x, y\}$. In particular, condition a) is fulfilled by construction, because s is randomized; condition b) is tested by condition one of theorem 1; and condition c) holds if all three conditions of theorem 1 are fulfilled. As such, the SCI algorithm may be used to screen neuroimaging data for pairs $\{x, y\}$ for which s qualifies as an IV. Under certain constraints, this enables the computation of the average causal effect of x on y [29, 44]. In the present work, we only use the SCI algorithm to test whether x is a cause of y.

3. Results

In this section, we study the performance of the SCI algorithm on simulated as well as on experimental data. We first investigate the power and the FDR of the SCI algorithm on simulated data that we generated with neural mass models for spectral responses in electrophysiology [30]. We then demonstrate the application of the SCI algorithm to a group-level causal analysis on experimental EEG data that we recorded as part of a study on brain-computer interfacing [31].

3.1. Simulation results

The power and the FDR of the SCI algorithm describe its ability to correctly detect a causal influence of x on y and to ignore causal models in which

this is not the case, respectively. Both metrics are likely to depend on various experimental parameters, such as the strength of the correlations between s, x, and y, the number of available i.i.d. samples of $\{s, x, y\}$, the significance level α_{rej} for rejecting independence, and the significance level α_{acc} for accepting independence. In this section, we estimate the power and the FDR of the SCI algorithm when varying these parameters. We base these estimates on simulation data that we generated with a neural mass model (NMM) for spectral responses in electrophysiology [30]. This model is an extension of the Jansen-Rit model [46]. It consists of a set of non-linear ordinary differential equations (ODEs) that model the dynamics of postsynaptic membrane potentials of excitatory and inhibitory cells in a cortical column. Solving the ODEs by numerical integration results in time-series which resemble the spectral characteristics of electromagnetic fields that are generated by cortical columns. Coupling multiple NMMs and varying their connection strengths then enables us to study the performance of the SCI algorithm in a variety of experimental settings.

We simulated the five causal models shown in figure 1. We did not include causal models in the simulations in which s is not a cause of x and of y, because such models can be rejected by traditional statistical methods [13]. For each of the DAGs in figure 1, the nodes x and y were simulated by one NMM, with the edges between the nodes specifying the connections between the NMMs. The stimulus s was modeled by applying a bandpass filter in the γ-range (55–85 Hz) to white noise and varying the noise amplitude in accordance with the stimulus labels (plus or minus one) around a common offset. For each of the five models, we simulated four experimental settings, in which we varied the strength of the stimulus' input to the NMMs to obtain different correlation strengths between s, x, and y. For each of the four settings, we simulated 500 trials (250 per stimulus label) with a duration of five seconds at a sampling frequency of 500 Hz. The generated time-series of each trial were then windowed with a Hann window and log-bandpower between 55 and 85 Hz was computed by a fast Fourier transform (FFT). This resulted, for each causal model and experimental setting, in a set of $n = 1, \ldots, 500$ i.i.d. samples $\{s_n, x_n, y_n\}$, where

s represents the stimulus' label (plus or minus one) and x and y represent trial-averaged log-bandpower in the γ-range of the two NMMs. For each causal model and experimental setting, we used all 500 trials to compute Pearson's correlation coefficients between s, x, and y, which are subsequently denoted as $\rho_{s,x}$, $\rho_{s,y}$, and $\rho_{x,y}$. Matlab code for generating the simulation data, including all parameters used in the simulations, is available as part of the SCI software package at `http://brain-computer-interfaces.net`.

To estimate the power and the FDR as a function of the number of available samples, we randomly drew N samples (without replacement) per stimulus label from each of the 20 models and then applied the SCI algorithm as described in table 1 and implemented in the SCI software package. We varied N between ten and 100 in steps of ten and repeated this process 500 times for each choice of N. We fixed $\alpha_{\text{rej}} = 0.01$ and varied α_{acc} between zero and 0.5 in steps of 0.01. For each setting, we then estimated the power of the SCI algorithm as the percentage of instances in which it correctly inferred a causal influence of x on y. We estimated the FDR as the percentage of instances in which the SCI algorithm incorrectly detected a causal influence of x on y.

The results of this simulation study are displayed in figure 2. The rows represent the causal models and the columns the experimental settings with various correlation strengths (displayed above each sub-figure). The colors depict the estimated power (for rows A and E) and FDR (for rows B–D) as a function of the significance level α_{acc} and of the number of samples per stimulus label. We note in row A that the power of the SCI algorithm increases with the number of available samples and with the strength of the correlations between the triple $\{s, x, y\}$. Its power decreases with α_{acc}. We note that the SCI algorithm requires strong correlations between $\{s, x, y\}$ and/or a large number of trials to achieve high power. For very strong correlations, the SCI algorithm achieves a power close to 100% with as few as 30 trials per condition (fourth and fifth column). We further note in rows B–D that with a reasonable number of samples per condition (≥ 30) and $\alpha_{\text{acc}} \geq 0.1$ the FDR remains below 2% in all experimental settings. Finally, we point out that, as predicted by theorem 1,

the SCI algorithm fails to detect a causal influence of x on y if s does not affect y solely via x (row E).

In summary, the SCI algorithm maintained a very low FDR across all tested simulation settings. We found its power to strongly depend on the strength of the correlations between $\{s, x, y\}$ and on the number of available samples. We suggest a choice of $\alpha_{\text{acc}} = 0.25$, $\alpha_{\text{rej}} = 0.01$ and at least 30 samples per condition. Based on the simulation results, we estimate this to enforce a FDR below 1.5% while resulting in a power of up to 90%. We emphasize that our simulation results indicate that the SCI algorithm requires reasonably large correlations between $\{s, x, y\}$ or a large number of samples to achieve high power. The power of the SCI algorithm can be enhanced by increasing α_{rej}, but this comes at the expense of a larger FDR.

3.2. Experimental results

We now demonstrate the application of the SCI algorithm to a group-level causal analysis on experimental data that we recorded as part of a study on brain-computer interfacing [31]. In this study, healthy subjects were trained via EEG-based neurofeedback to self-regulate the amplitude of γ-oscillations (55–85 Hz) in the right superior parietal cortex (SPC). The data from this study is particularly suited to illustrate the performance of the SCI algorithm, because, first, the right SPC is a primary node of the central executive network (CEN) [47], and, second, stimulation of the CEN by TMS has been found to modulate the medial prefrontal cortex (MPC) [33]. This allowed us to formulate the following causal hypothesis: Variations of γ-power in the right SPC (γ_{SPC}), which are induced by the instruction (s) to up- or down-regulate γ_{SPC}, modulate γ-power in the MPC (γ_{MPC}), i.e. $s \to \gamma_{\text{SPC}} \to \gamma_{\text{MPC}}$. We now demonstrate how to test this causal hypothesis with the SCI algorithm.

3.2.1. Experimental setup and data

Subjects were placed in front of a computer screen, about 1.5 m away, and received visual feedback on their current level of γ-power (55–85 Hz) in the right

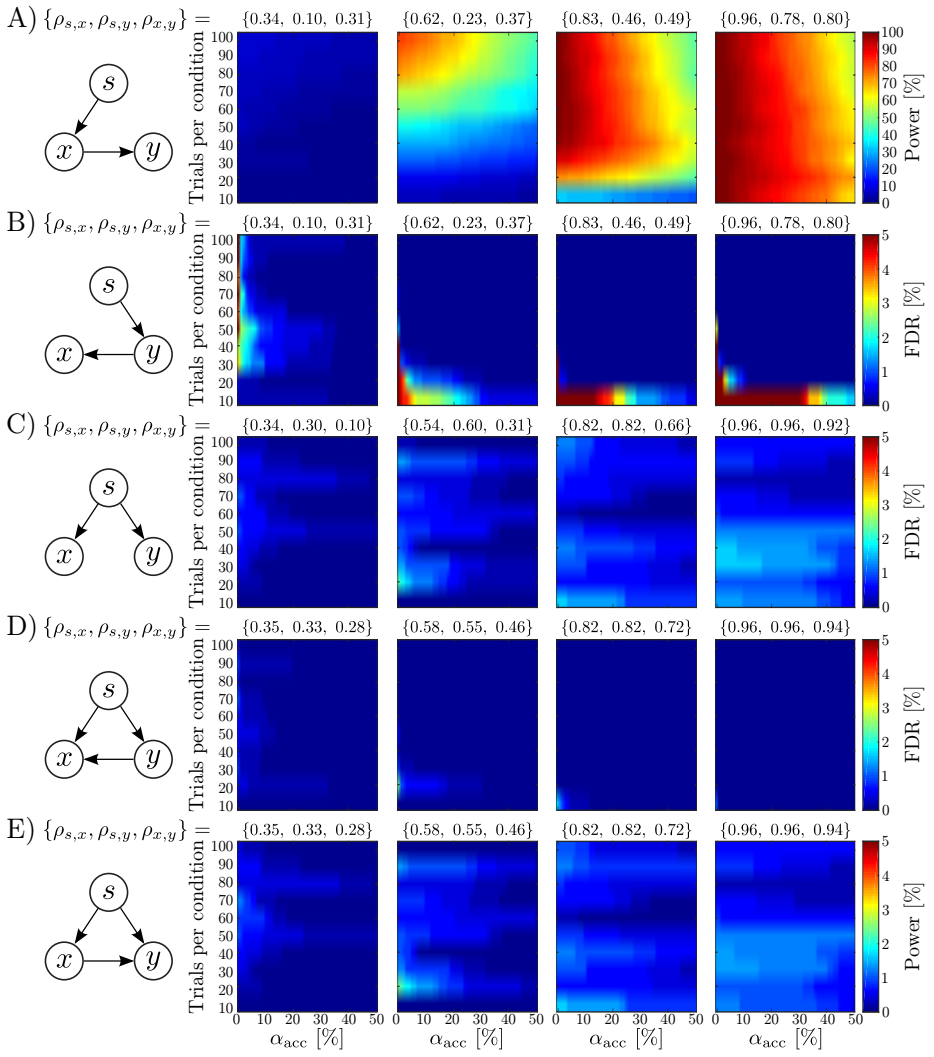

Figure 2: Power and FDR of the SCI algorithm on simulated spectral EEG responses as a function of the number of available samples and of the significance level α_{acc} for accepting independence. The significance level for rejecting independence was set to $\alpha_{\text{rej}} = 0.01$. The rows represent the simulated causal models. The columns represent experimental settings with varying correlation strengths. Note the different color range between row A and rows B–E.

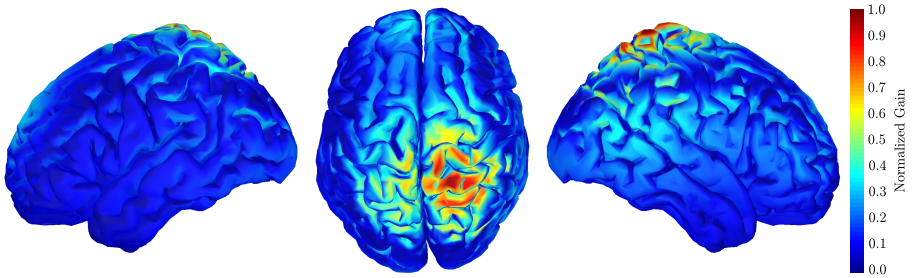

Figure 3: Spatial transfer function of the beamformer used for neurofeedback, focused on the right SPC (cf. [31] for details).

SPC. The feedback signal was computed by first applying a linearly-constrained-minimum-variance (LCMV) beamformer [48], aimed at the right SPC, to a 121-channel EEG recording at sampling rate of 500 Hz. The noise covariance matrix for the beamforming procedure was estimated from five-minute resting-state EEG data recorded prior to each feedback session. Figure 3 shows the spatial transfer function of the beamformer, focusing on the right SPC (cf. [31]). The past five seconds of the beamformed signal were then windowed with a Hann window, and the log-bandpower from 55 to 85 Hz was computed by a FFT. Then the current level of γ-power in the right SPC, subsequently termed γ_{SPC}, was linearly mapped to the vertical position of a white ball displayed on the computer screen. The center of the screen represented the median γ_{SPC} during the resting-state baseline, and the upper and lower limits of the screen represented the addition and subtraction of two standard deviations, respectively. The horizontal position of the ball was fixed to the center of the screen. Visual feedback was updated at a frame rate of 25 Hz. All online data processing was implemented in BCI2000 [49].

In each trial, subjects were instructed to move the feedback ball to a yellow rectangle displayed at the top or bottom of the screen. After one minute the white ball disappeared, and the subjects were instructed to rest for five seconds before attempting the next trial. In each session, subjects performed three blocks of 20 trials in pseudo-randomized order with a brief intermission between

each block, giving a total of 30 trials per condition.

Three healthy subjects were invited for five feedback sessions each. Because the subjects required multiple training sessions to reliably self-regulate γ_{SPC}, we only analyze data from the last two sessions of each subject.

3.2.2. Data pre-processing

To attenuate confounding by non-cortical processes, we cleaned the data of each session of artifacts by using an independent component analysis (ICA) [50]. Specifically, we first high-pass filtered each data set by using a 3rd-order Butterworth filter with a cut-off frequency of 3 Hz, reduced it to 64 dimensions by principal component analysis (PCA), and then separated it into independent components (ICs) using the SOBI algorithm [51]. We then sorted ICs according to their neurophysiological plausibility [52], and inspected each of their topography, spectrum, and time-series. We rejected an IC as non-cortical when it exhibited at least one of the following four criteria [53, 31]: (1) The spectrum did not show the $1/f$-behavior typical of a cortical source. In particular, we rejected ICs that displayed a monotonic increase in spectral power starting around 20 Hz, which is characteristic of muscular activity [54]. (2) Eye blinks were detectable in the time-series. (3) The topography did not show a dipolar pattern. (4) The time-series appeared to be contaminated by other noise sources such as 50 Hz line noise or large spikes. We then re-projected the remaining ICs to the scalp. Note that, while ICA can attenuate artifacts in the signal, the ill-posed nature of the EEG inverse problem means that non-cortical processes cannot be completely eliminated [55].

To investigate causal relations between cortical sources, rather than between EEG electrodes, we performed a source-localization procedure. We used an LCMV beamformer on the raw EEG data of each session to estimate the time courses of $M = 15028$ current dipoles distributed across the cortical surface, using standardized electrode locations and a three-shell spherical head model [56]. We further applied the beamformer, which we used for the online neurofeedback in each session, to compute the time course of the neurofeedback signal after at-

tenuating non-cortical processes by ICA. We then computed the trial-averaged log-bandpower in the γ-range of each signal, again using a FFT in conjunction with a Hann window. Thus, for each recording session, we obtained $N = 60$ samples of the M data sets $\mathcal{V}_n^m = \{s_n,\ \gamma_{\mathrm{SPC},n},\ \gamma_{m,n}\}$; s_n represents the visual stimulus in the n^{th} trial (coded as plus or minus one) that instructs subjects to either up- or down-regulate γ_{SPC}, $\gamma_{\mathrm{SPC,n}}$ is the trial-averaged log-bandpower in the γ-range in the right SPC, and $\gamma_{m,n}$ is the trial-averaged log-bandpower in the γ-range at the m^{th} cortical location.

We note that the mean correlation strengths across all subjects, sessions and cortical locations for this data set are $\rho_{s,\gamma_{\mathrm{SPC}}} = 0.68$, $\rho_{s,\gamma_m} = 0.61$, and $\rho_{\gamma_{\mathrm{SPC}},\gamma_n} = 0.81$. Based on the simulation results in section 3.1, these correlation strengths are sufficient to expect a high power of the SCI algorithm.

We further note that while beamforming is known to be blind to perfectly correlated sources [48, 57], such sources violate the faithfulness assumption (condition three of theorem 1 is trivially true when x and y are perfectly correlated). We thus consider this property of beamformers an advantageous feature in the present context.

3.2.3. Group-level SCI results

We used the SCI algorithm to test the causal hypothesis $s \rightarrow \gamma_{\mathrm{SPC}} \rightarrow \gamma_m$ for each of the cortical sources estimated by the source localization procedure, as described in section 3.2.2. To apply the SCI algorithm to group-level data, while avoiding problems associated with pooling data from multiple subjects [9], we computed a group-level statistic. Specifically, we used the fact that, by definition, p-values are drawn from a uniform distribution with support from zero to one if the null-hypothesis is true. We will now illustrate how this let us test null-hypotheses on the group level on $H0_1 : s \perp\!\!\!\perp \gamma_{\mathrm{SPC}}$, the null-hypothesis which states that the instruction given to subjects in each trial is independent of γ-power in the right SPC (cf. step 3 in table 1).

We first used a correlation analysis to test this null-hypothesis on the level of individual sessions, as described in section 2.4. Table 2 shows the correlation

Table 2: Pearson's correlation coefficients between γ_{SPC} and the experimental conditions (up- and down-regulation of γ_{SPC}).

	Subject 1	Subject 2	Subject 3
Session 4	0.80 ($p < 10^{-4}$)	0.35 ($p = 0.0026$)	0.86 ($p < 10^{-4}$)
Session 5	0.84 ($p < 10^{-4}$)	0.37 ($p = 0.0021$)	0.83 ($p < 10^{-4}$)

coefficients and associated p-values generated by this analysis. To test $H0_1$ on the group level, we then computed the empirical cumulative distribution function (CDF) of these p-values and quantified its deviation from the CDF of a uniform distribution of p-values. Specifically, we created one hundred bins between zero and one and summed, across all bins, the absolute differences between the empirically observed CDF and the one generated by drawing the same number of samples from a uniform distribution between zero and one. We then sampled this test statistic from the null-distribution 10^4 times. This let us estimate the probability of observing the p-values in table 2 under $H0_1$. We found $p < 10^{-4} < \alpha_{\mathrm{rej}} = 0.01$ for the resulting group-level p-value, so we rejected $H0_1$ on the group level.

We then tested $H0_2 : s \perp\!\!\!\perp \gamma_m$ on the group level in the same manner. Because there are multiple comparisons for each of the M cortical sources, we corrected the group-level significance level using a FDR of 0.01 [58].

Finally, we performed a linear regression for each of the M cortical sources from γ_{SPC} to γ_m to estimate the residuals $\hat{\gamma}_m$, and then tested $H0_3 : \hat{\gamma}_m \perp\!\!\!\perp (s, \gamma_{\mathrm{SPC}})$ using the HSIC criterion described in section 2.4. Using the same group-level approach as for $H0_1$ and $H0_2$, but without using FDR correction to remain conservative, we accepted $H0_3$ if we found the resulting group-level p-value to be greater than $\alpha_{\mathrm{acc}} = 0.25$.

Figure 4 shows a cortical map of sources for which we rejected $H0_1$ and $H0_2$ but accepted $H0_3$; that is, the cortical sources which we inferred to be modulated by γ-power in the right SPC. Consistent with our hypothesis, we found the most prominent modulation target of the right SPC to be bilaterally

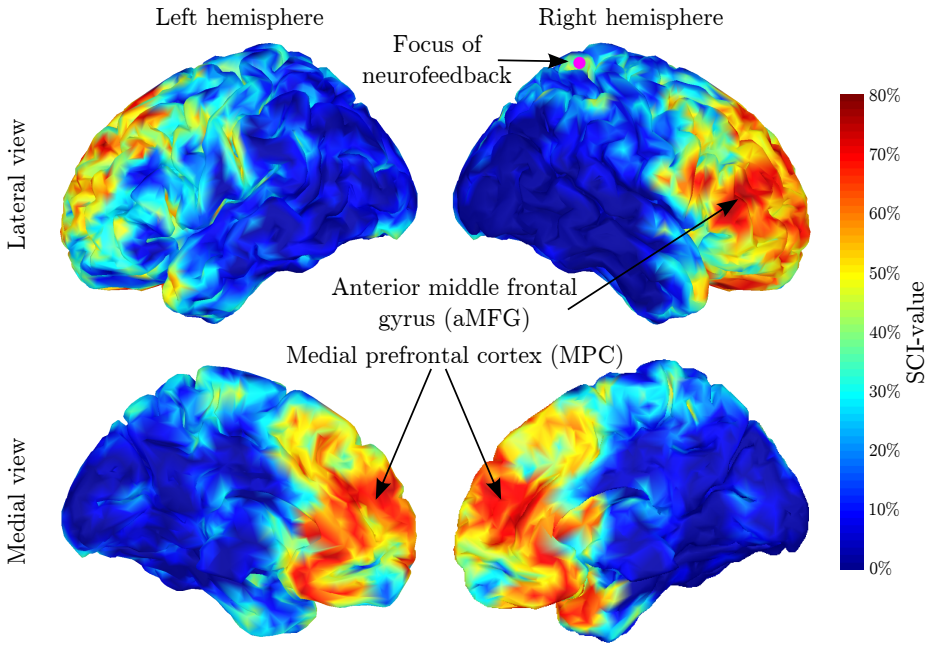

Figure 4: Results of the SCI analysis. The SCI value denotes the percentage of dipoles within a radius of 7 mm that are modulated by the neurofeedback target area in the right SPC, indicated by a pink circle. From these results, we infer the primary targets of the right SPC to be the MPC and the right aMFG.

in the MPC. In addition, the SCI algorithm inferred that the right anterior middle frontal gyrus (aMFG), a node of the salience network [59, 33], was also modulated by the right SPC.

To provide further empirical support for these conclusions, we varied α_{acc} between 0.1 and 0.8. This did not have a qualitative effect on our causal conclusions. Furthermore, we repeated the SCI analysis while reversing the direction of the causal tests; that is, we tested whether $s \to \gamma_m \to \gamma_{\text{SPC}}$ for each of the M dipoles. For the original choice of $\alpha_{\text{acc}} = 0.25$, this reversal caused the analysis to generate only $9/15028 = 0.06\%$ positive test results scattered throughout the brain, relative to $4674/15028 = 31.1\%$ positive tests in the original causal direction.

4. Discussion

We conclude the article with a discussion of the utility of causal inference in neuroimaging (section 4.1), the strengths and (current) limitations of the SCI algorithm (section 4.2), and the promise the SCI algorithm holds for the study of the neural basis of cognition (4.3).

4.1. Causal inference in neuroimaging

Only interventional studies can prove causal relationships. Interventional studies, however, are often costly and time-consuming. Causal inference methods can be used to screen large datasets for potential causal relations, which may then be validated by subsequent interventional studies. In this way, the number of interventional studies, that are on average required to prove a single causal relation, can be substantially reduced [60]. In order to exploit this strength, it is essential that causal inference methods predict the effects of external manipulations. Methods based on CBNs, including the SCI algorithm, give testable predictions on the effects of external manipulations [18], while it is less straightforward to do so in frameworks based on information flow [35, 61]. We consider it essential for progress in neuroimaging that studies on brain connectivity clearly distinguish between these concepts.

4.2. Summary of the SCI algorithm

The utility of causal inference for planning future interventional studies depends on the power and on the FDR of the employed method. In order to reduce the probability of false positive results, it is essential to control for latent confounding. The primary strength of the SCI algorithm is its ability to control for any latent confounder. Its further advantages are, first, that it does not require brain state features to have a jointly Gaussian distribution [21], and, second, that it implicitly tests its inherent assumptions. Because theorem 2 provides sufficient conditions for conditional independence, the presence of empirical evidence suggesting a genuine causal influence means that the linearity assumption of the regression-based conditional independence test is fulfilled. The drawbacks

of our method are, first, that it is only applicable in stimulus-based settings, and, second, that it fails to find genuine causal relations in the presence of non-linear dependencies. We do not consider the latter issue a strong limitation, however, because the empirical evidence for BOLD signals and electrophysiological recordings suggests that their feature relations are predominantly linear [62, 63]. It implies, however, that a lack of evidence in favor of a genuine causal relation may not be interpreted as evidence against it.

The assumption of faithfulness, however, is difficult to test on empirical data [64]. Faithfulness states that all (conditional) independence relations in observed data are implied by the generating DAG. In theory, faithfulness is a rather weak assumption. Because the set of unfaithful distributions for a given DAG has measure zero [65], it is a priori unlikely that experimental data is sampled from an unfaithful distribution. In practice, however, distributions close to unfaithfulness may make it difficult to discover residual dependencies in finite data, which may generate false positive test results [66]. While our simulation results suggest that the SCI algorithm has a low FDR in a variety of experimental settings, its performance may be further optimized by considering different kernels and/or optimizing the bandwidth for the HSIC test.

4.3. Causal inference on the neural basis of cognition

In the present article, we have focused on studying causal relations between pairs of neural states. The SCI algorithm, however, can be applied to any arbitrary combination of data modalities. As such, it is straightforward to apply it to the study of the neural basis of cognition. Specifically, we can test causal hypotheses of the form $s \to x \to r$, where s represents an experimental stimulus, x refers to a neural state, and r measures a behavioral response. In this way, we can ask whether a brain state x is likely to be a cause of a behavioral response r. We note that r may operationalize cognitive states of arbitrary complexity, ranging from simple behavioral responses to complex personality traits. It may not be trivial, however, to identify experimental stimuli that modulate a certain type of response variable. Depending on the complexity of r, it may also not

be sufficient to consider one-dimensional neural states x. Instead, we may have to consider causal models of the form $s \rightarrow \boldsymbol{w}^{\mathrm{T}}\boldsymbol{x} \rightarrow r$, where $\boldsymbol{x}$ is a vector of neural states and we wish to identify the linear combination(s) $\boldsymbol{w}$ such that the conditions tested by the SCI algorithm are fulfilled. This would enable us to consider multivariate patterns of brain activity as the neural bases for cognitive states.

References

[1] S. Smith, K. Miller, G. Salimi-Khorshidi, M. Webster, C. Beckmann, T. Nichols, J. Ramsey, M. Woolrich, Network modelling methods for fMRI, NeuroImage 54 (2) (2011) 875–891.

[2] A. McIntosh, F. Gonzalez-Lima, Structural equation modeling and its application to network analysis in functional brain imaging, Human Brain Mapping 2 (1-2) (1994) 2–22.

[3] L. Atlas, N. Bolger, M. Lindquist, T. Wager, Brain mediators of predictive cue effects on perceived pain, The Journal of Neuroscience 30 (39) (2010) 12964–12977.

[4] C. Granger, Investigating causal relations by econometric models and cross-spectral methods, Econometrica: Journal of the Econometric Society 37 (3) (1969) 424–438.

[5] M. Kamiński, M. Ding, W. Truccolo, S. Bressler, Evaluating causal relations in neural systems: Granger causality, directed transfer function and statistical assessment of significance, Biological Cybernetics 85 (2) (2001) 145–157.

[6] G. Gregoriou, S. Gotts, H. Zhou, R. Desimone, High-frequency, long-range coupling between prefrontal and visual cortex during attention, Science 324 (5931) (2009) 1207–1210.

[7] K. Friston, L. Harrison, W. Penny, Dynamic causal modelling, NeuroImage 19 (4) (2003) 1273–1302.

[8] J. Daunizeau, O. David, K. Stephan, Dynamic causal modelling: a critical review of the biophysical and statistical foundations, NeuroImage 58 (2) (2011) 312–322.

[9] J. Ramsey, S. Hanson, C. Hanson, Y. Halchenko, R. Poldrack, C. Glymour, Six problems for causal inference from fMRI, NeuroImage 49 (2) (2010) 1545–1558.

[10] M. Grosse-Wentrup, B. Schölkopf, J. Hill, Causal influence of gamma oscillations on the sensorimotor rhythm, NeuroImage 56 (2) (2011) 837–842.

[11] J. Ramsey, S. Hanson, C. Glymour, Multi-subject search correctly identifies causal connections and most causal directions in the DCM models of the smith et al. simulation study, NeuroImage 58 (3) (2011) 838–848.

[12] J. Mumford, J. Ramsey, Bayesian networks for fMRI: A primer, Neuroimage 86 (2014) 573–582.

[13] S. Weichwald, T. Meyer, O. Özdenizci, B. Schölkopf, T. Ball, M. Grosse-Wentrup, Causal interpretation rules for encoding and decoding models in neuroimaging, NeuroImage 110 (2015) 48–59.

[14] K. Friston, B. Li, J. Daunizeau, K. Stephan, Network discovery with DCM, NeuroImage 56 (3) (2011) 1202–1221.

[15] M. Kern, A. Aertsen, A. Schulze-Bonhage, T. Ball, Heart cycle-related effects on event-related potentials, spectral power changes, and connectivity patterns in the human ECoG, Neuroimage 81 (2013) 178–190.

[16] M. Grosse-Wentrup, Understanding brain connectivity patterns during motor imagery for brain-computer interfacing, in: D. Koller, D. Schuurmans, Y. Bengio, L. Bottou (Eds.), Advances in Neural Information Processing Systems 21, 2009, pp. 561–568.

[17] J. Hipp, M. Siegel, Dissociating neuronal gamma-band activity from cranial and ocular muscle activity in EEG, Frontiers in Human Neuroscience 7 (338) (2013) 1–11. doi:10.3389/fnhum.2013.00338.

[18] J. Pearl, Causality: Models, Reasoning, and Inference, Cambridge University Press, 2000.

[19] P. Spirtes, C. Glymour, R. Scheines, Causation, Prediction, and Search, MIT Press, 2000.

[20] J. Zhang, On the completeness of orientation rules for causal discovery in the presence of latent confounders and selection bias, Artificial Intelligence 172 (16) (2008) 1873–1896.

[21] L. Waldorp, I. Christoffels, V. van de Ven, Effective connectivity of fMRI data using ancestral graph theory: dealing with missing regions, NeuroImage 54 (4) (2011) 2695–2705.

[22] S. Hanson, B. Bly, The distribution of BOLD susceptibility effects in the brain is non-Gaussian, NeuroReport 12 (9) (2001) 1971–1977.

[23] A. Wink, J. Roerdink, BOLD noise assumptions in fMRI, International Journal of Biomedical Imaging 2006.

[24] A. Gretton, R. Herbrich, A. Smola, O. Bousquet, B. Schölkopf, Kernel methods for measuring independence, The Journal of Machine Learning Research 6 (2005) 2075–2129.

[25] A. Gretton, K. Fukumizu, C. Teo, L. Song, B. Schölkopf, A. Smola, A kernel statistical test for independence, in: J. Platt, D. Koller, Y. Singer, S. Roweis (Eds.), Advances in Neural Information Processing Systems 20, MIT Press, 2008, pp. 585–592.

[26] A. Gretton, L. Györfi, Consistent nonparametric tests of independence, The Journal of Machine Learning Research 11 (2010) 1391–1423.

[27] K. Fukumizu, A. Gretton, X. Sun, B. Schölkopf, Kernel measures of conditional dependence, in: J. Platt, D. Koller, Y. Singer, S. Roweis (Eds.), Advances in Neural Information Processing Systems 20, MIT Press, Cambridge, MA, 2008, pp. 489–496.

[28] K. Zhang, J. Peters, D. Janzing, B. Schölkopf, Kernel-based conditional independence test and application in causal discovery, in: Proceedings of the 27th Conference Annual Conference on Uncertainty in Artificial Intelligence (UAI 2011), AUAI Press, Corvallis, OR, 2011, pp. 804–813.

[29] J. Angrist, G. Imbens, D. Rubin, Identification of causal effects using instrumental variables, Journal of the American Statistical Association 91 (434) (1996) 444–455.

[30] R. Moran, S. Kiebel, K. Stephan, R. Reilly, J. Daunizeau, K. Friston, A neural mass model of spectral responses in electrophysiology, NeuroImage 37 (3) (2007) 706–720.

[31] M. Grosse-Wentrup, B. Schölkopf, A brain–computer interface based on self-regulation of gamma-oscillations in the superior parietal cortex, Journal of Neural Engineering 11 (5) (2014) 056015.

[32] S. Bressler, V. Menon, Large-scale brain networks in cognition: emerging methods and principles, Trends in Cognitive Sciences 14 (6) (2010) 277–290.

[33] A. Chen, C. Oathes, D.J.and Chang, T. Bradley, Z.-W. Zhou, L. Williams, K. Glover, G.H.and Deisseroth, A. Etkin, Causal interactions between fronto-parietal central executive and default-mode networks in humans, Proceedings of the National Academy of Sciences 110 (49) (2013) 19944–19949.

[34] A. Roebroeck, E. Formisano, R. Goebel, Mapping directed influence over the brain using granger causality and fMRI, Neuroimage 25 (1) (2005) 230–242.

[35] J. Lizier, M. Prokopenko, Differentiating information transfer and causal effect, The European Physical Journal B-Condensed Matter and Complex Systems 73 (4) (2010) 605–615.

[36] P. Holland, Statistics and causal inference, Journal of the American Statistical Association 81 (396) (1986) 945–960.

[37] P. Bühlmann, M. Kalisch, M. Maathuis, Variable selection in high-dimensional linear models: partially faithful distributions and the PC-simple algorithm, Biometrika 97 (2) (2010) 261–278.

[38] K. Friston, A. Holmes, K. Worsley, J. Poline, C. Frith, R. Frackowiak, Statistical parametric maps in functional imaging: a general linear approach, Human brain mapping 2 (4) (1994) 189–210.

[39] A. Delorme, S. Makeig, EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis, Journal of Neuroscience Methods 134 (1) (2004) 9–21.

[40] I. Steinwart, On the influence of the kernel on the consistency of support vector machines, The Journal of Machine Learning Research 2 (2002) 67–93.

[41] A. Gretton, D. Sejdinovic, H. Strathmann, S. Balakrishnan, M. Pontil, K. Fukumizu, B. Sriperumbudur, Optimal kernel choice for large-scale two-sample tests, in: Advances in Neural Information Processing Systems, 2012, pp. 1205–1213.

[42] R. Bowden, S. Turkington, Instrumental Variables, Cambridge University Press, 1984.

[43] M. Arellano, O. Bover, Another look at the instrumental variable estimation of error-components models, Journal of Econometrics 68 (1) (1995) 29–51.

[44] V. Didelez, N. Sheehan, Mendelian randomization as an instrumental variable approach to causal inference, Statistical Methods in Medical Research 16 (4) (2007) 309–330.

[45] L. Chen, F. Emmert-Streib, J. Storey, Harnessing naturally randomized transcription to infer regulatory relationships among genes, Genome Biology 8 (10) (2007) R219.

[46] B. Jansen, V. Rit, Electroencephalogram and visual evoked potential generation in a mathematical model of coupled cortical columns, Biological Cybernetics 73 (4) (1995) 357–366.

[47] V. Menon, Large-scale brain networks and psychopathology: a unifying triple network model, Trends in Cognitive Sciences 15 (10) (2011) 483–506.

[48] B. van Veen, W. van Drongelen, M. Yuchtman, A. Suzuki, Localization of brain electrical activity via linearly constrained minimum variance spatial filtering, IEEE Transactions on Biomedical Engineering 44 (1997) 867–880.

[49] G. Schalk, D. McFarland, T. Hinterberger, N. Birbaumer, J. Wolpaw, BCI2000: A general-purpose brain-computer interface (BCI) system, IEEE Transactions on Biomedical Engineering 51 (6) (2004) 1034–1043.

[50] A. Delorme, T. Sejnowski, S. Makeig, Enhanced detection of artifacts in EEG data using higher-order statistics and independent component analysis, Neuroimage 34 (4) (2007) 1443–1449.

[51] A. Belouchrani, K. Abed-Meraim, J.-F. Cardoso, E. Moulines, A blind source separation technique using second-order statistics, IEEE Transactions on Signal Processing 45 (2) (1997) 434–444.

[52] M. Grosse-Wentrup, B. Schölkopf, A review of performance variations in SMR-based brain-computer interfaces (BCIs), in: C. Guger, B. Allison, G. Edlinger (Eds.), Brain-Computer Interface Research, Springer, 2013, pp. 39–51.

[53] M. Grosse-Wentrup, B. Schölkopf, High gamma-power predicts performance in sensorimotor-rhythm brain-computer interfaces, Journal of Neural Engineering 55 (2012) 1991–2000.

[54] I. Goncharova, D. McFarland, T. Vaughan, J. Wolpaw, EMG contamination of EEG: Spectral and topographical characteristics, Clinical Neurophysiology 14 (2003) 1580–1593.

[55] B. McMenamin, A. Shackman, J. Maxwell, D. Bachhuber, A. Koppenhaver, L. Greischar, R. Davidson, Validation of ICA-based myogenic artifact correction for scalp and source-localized EEG, NeuroImage 49 (2010) 2416–2432.

[56] J. Mosher, S. Baillet, F. Darvas, D. Pantazis, E. Yildirim, R. Leahy, Brainstorm electromagnetic imaging software, in: 5th International Symposium on Noninvasive Functional Source Imaging within the Human Brain and Heart (NFSI 2005), 2005.

[57] K. Sekihara, S. Nagarajan, D. Poeppel, A. Marantz, Performance of an MEG adaptive-beamformer technique in the presence of correlated neural activities: effects on signal intensity and time-course estimates, IEEE Transactions on Biomedical Engineering 49 (12) (2002) 1534–1546.

[58] Y. Benjamini, Y. Hochberg, Controlling the false discovery rate: a practical and powerful approach to multiple testing, Journal of the Royal Statistical Society. Series B (Methodological) (1995) 289–300.

[59] W. Seeley, V. Menon, A. Schatzberg, J. Keller, G. Glover, H. Kenna, A. Reiss, M. Greicius, Dissociable intrinsic connectivity networks for salience processing and executive control, The Journal of Neuroscience 27 (9) (2007) 2349–2356.

[60] M. Maathuis, D. Colombo, M. Kalisch, P. Bühlmann, Predicting causal effects in large-scale systems from observational data, Nature Methods 7 (4) (2010) 247–248.

[61] M. Eichler, V. Didelez, On Granger causality and the effect of interventions in time series, Lifetime Data Analysis 16 (1) (2010) 3–32.

[62] K. Müller, C. W. Anderson, G. E. Birch, Linear and nonlinear methods for brain-computer interfaces, IEEE Transactions on Neural Systems and Rehabilitation Engineering 11 (2) (2003) 165–169.

[63] T. Naselaris, K. Kay, S. Nishimoto, J. Gallant, Encoding and decoding in fMRI, Neuroimage 56 (2) (2011) 400–410.

[64] J. Zhang, P. Spirtes, Strong faithfulness and uniform consistency in causal inference, in: Proceedings of the nineteenth conference on uncertainty in artificial intelligence, Morgan Kaufmann Publishers Inc., 2002, pp. 632–639.

[65] C. Meek, Strong completeness and faithfulness in Bayesian networks, in: Proceedings of the Eleventh Conference Annual Conference on Uncertainty in Artificial Intelligence (UAI-95), Morgan Kaufmann, San Francisco, CA, 1995, pp. 411–418.

[66] C. Uhler, G. Raskutti, P. Bühlmann, B. Yu, Geometry of the faithfulness assumption in causal inference, The Annals of Statistics 41 (2) (2013) 436–463.