Robust inference of causality in high-dimensional dynamical processes from the Information Imbalance of distance ranks

Vittorio Del Tatto[a] (ID), Gianfranco Fortunato[a] (ID), Domenica Bueti[a] (ID), and Alessandro Laio[a,b,1]

We introduce an approach which allows detecting causal relationships between variables for which the time evolution is available. Causality is assessed by a variational scheme based on the Information Imbalance of distance ranks, a statistical test capable of inferring the relative information content of different distance measures. We test whether the predictability of a putative driven system Y can be improved by incorporating information from a potential driver system X, without explicitly modeling the underlying dynamics and without the need to compute probability densities of the dynamic variables. This framework makes causality detection possible even between high-dimensional systems where only few of the variables are known or measured. Benchmark tests on coupled chaotic dynamical systems demonstrate that our approach outperforms other model-free causality detection methods, successfully handling both unidirectional and bidirectional couplings. We also show that the method can be used to robustly detect causality in human electroencephalography data.

time series analysis | causality | dynamical systems

Discovering causal relationships among observable quantities has been inspiring and guiding scientific research from its dawn, as causality is at the very heart of physical phenomena and natural laws. The definition of causality is far from univocal, with diverse frameworks rooted in distinct perspectives. Granger's paradigm is based on "predictive causality" (1), while Pearl's structural causal model (2) builds on counterfactuals. Determining causality from data collected without directly intervening on the system under study—namely, without performing interventional experiments where the causal variable is manipulated—is a challenging problem which received increasing attention over the last decades (3–5). The use of purely observational data is the only option when experiments are unfeasible or unethical, such as in the case of medical studies that would create a real risk for patient's health, or Earth science research that could alter delicate ecological balances, just to give a few examples. This motivated the development of statistical tests aimed specifically at inferring causal relationships in "real world" time-ordered data. These approaches are routinely employed in diverse fields, from economics (1, 6) to ecology (7), Earth system sciences (8) and neuroscience (9–11). The common idea to all these methods is to compute statistical measures which are asymmetric under the exchange of the dynamic variables, in order to reflect the asymmetry of a putative causal coupling. In this field, some important conceptual and practical problems remain open and are still object of intense investigation. In particular, the fact that real-world time series often emerge from complex underlying dynamics naturally brings to the necessity of methodologies dealing with high-dimensional data (12, 13). Moreover, false-positive detections represent a common yet crucial limitation even in low-dimensional scenarios (14, 15).

From a historical perspective, the first quantitative criterion to measure causality dates back to the work of Wiener (16), who postulated that the prediction of a signal Y can be improved by using the past information of a signal X if X is causal to Y. Inspired by Wiener, Granger proposed to identify causal links in time series analysis with a vector autoregressive model assuming a linear dynamics (1). Since then, several nonlinear generalizations of Granger's idea have been proposed (17–22). In particular, the Extended Granger Causality test (17) confines the linear approximation of the dynamics to local regions in the state space. Causality can also be inferred by estimating a conditional mutual information named Transfer Entropy (23–25), which is equivalent to Granger causality for Gaussian variables (26). However, computing multivariate probability distributions to evaluate mutual information is challenging in high-dimensional systems, and in practice one is typically forced to work with conditional probabilities of a few variables at a time (27). For this reason, alternative methods that do not require computing the

Significance

Uncovering causality from purely observational time-dependent data is a problem with countless applications. Many available methods are often not able to distinguish the absence of causality from a weak causal link, bringing to false-positive discovery. We introduce a method which strongly mitigates this problem. Importantly, the approach does not require assuming a specific form for the dynamics generating the data. Compared to other available methods, we find that both on synthetic and real-world data, our approach detects a considerably smaller fraction of false positives, allowing a more rigorous identification of the variables which determine the future evolution of a system.

Author affiliations: [a]Physics Section, Scuola Internazionale Superiore di Studi Avanzati, Trieste 34136, Italy; and [b]Condensed Matter and Statistical Physics Section, International Centre for Theoretical Physics, Trieste 34151, Italy

https://doi.org/10.1073/pnas.2317256121   **1 of 10**

probability distributions of the dynamic variables are more appealing for real-world applications. Among these, cross-mapping methods rely on Takens' theorem (28), which allows reconstructing a dynamical system's attractor—or rather a version capturing its main features, called shadow manifold—using one-dimensional time series. In particular, Convergent Cross-Mapping (7) evaluates the coupling strength $X \rightarrow Y$ by attempting a local reconstruction of the shadow manifold of X from the shadow manifold of Y and computing a correlation coefficient between the reconstructed and the ground-truth points. Another cross-mapping method, known as measure L (29), employs a similar approach, but carries out the local reconstruction of the target manifold using ranks rather than distances.

In this work, we introduce a causality detection method broadly based on Granger causality principles, namely that the cause occurs before the effect and the cause contains unique information about the effect. We implement these principles using the Information Imbalance (30), a statistical measure designed to compare distance spaces and decide which space is more informative, without modeling the underlying dynamics. By measuring distances between independent realizations of the same dynamical process, the method evaluates whether including the putative driver variables in the distance space built at time $t = 0$ allows to better guess which pairs of trajectories will be the closest in the space of the driven variables at a future time $t = \tau$. The information of the driver system is added to the space of the putative driven system using a variational approach, which allows probing the presence or absence of a coupling within a theoretically rigorous framework. The distances in different spaces are compared by analyzing the statistics of distance ranks and examining how it is affected by the inclusion of potential causal variables in the distance definition. Ranks can be easily obtained by ordering distances from the smallest to the largest, and the effort to compute them is not affected by the underlying dimensionality.

To assess the validity of the method we carried out tests on a variety of coupled dynamical systems with both unidirectional and bidirectional couplings, as well as on real-world time series from electroencephalography (EEG) experiments. These tests led us to observe that our method allows recognizing with high statistical confidence when a causal coupling is absent. Remarkably, we find that the other approaches we tested fail systematically in this task, bringing to high probabilities of observing false positives, namely of confusing the absence of causality with a condition of weak causal coupling. Our approach, besides strongly mitigating this problem, provides reliable results also when the dynamics system is high-dimensional, such as the electrophysiological signal of a human brain, or a Lorenz 96 system (31).

The Information Imbalance Gain

The Information Imbalance is a statistical measure introduced to compare the information content of two distances d_A and d_B defined on a set of points $\{x_i\}$ $(i = 1, ..., N)$, which allows assessing whether the distances are equivalent, independent or if one is more informative than the other (30).

This measure is built on the idea that close points according to d_A remain close in distance space d_B when d_A is informative with respect to d_B or, equivalently, when the information carried by d_B is also contained in d_A. We denote by r_{ij}^A (resp. r_{ij}^B) the rank of point j with respect to point i according to distance d_A (resp. d_B),

with the convention $r_{ii}^A = r_{jj}^B = 0$. For example, $r_{ij}^A = 1$ if j is the nearest neighbor of i in space A. The Information Imbalance from A to B is defined as the average rank according to distance d_B restricted to points which are "close" according to d_A:

$$\Delta(d_A \rightarrow d_B) = \frac{2}{N} \langle r^B \mid r^A \le k \rangle = \frac{2}{N^2 k} \sum_{\substack{i,j \\ \text{s.t. } r_{ij}^A \le k}} r_{ij}^B. \quad [1]$$

The parameter k specifies the number of neighbors taken into account and generalizes the definition in ref. 30, which assumes $k = 1$. This definition can also be interpreted as an asymptotic upper bound of a restricted mutual information, that we define in the *SI Appendix*, section 1. The prefactor $2/N$ statistically confines this quantity between 0 and 1, which are the limits of d_A being respectively maximally and minimally informative with respect to d_B. As discussed in the *SI Appendix*, section 2A, in this form the Information Imbalance is equivalent to the measure L introduced by Chicharro and Andrzejak in ref. 29 for studying coupled dynamical systems. However, the way we apply this statistic to the problem of causal detection is substantially different, as discussed below.

We here propose a variational approach, based on the Information Imbalance, to infer the presence of causal relationships among two sets of time-dependent variables, using no prior knowledge about the underlying dynamics. To set the framework, let $X(t)$ and $Y(t)$ be vectors characterizing the states of two dynamical systems at time t, with components $x_\alpha(t)$ and $y_\beta(t)$ $(\alpha = 1, ..., D; \beta = 1, ..., D')$. We suppose that all the components of X (and, separately, of Y) are dynamically intertwined, namely that there are no proper subsets of coordinates of X (resp. of Y) that are autonomous. This condition implies that the components of X and Y cannot be regrouped in three distinct systems X', Y', and Z' such that Z' is unidirectionally coupled both to X' and Y', namely a so-called common driver (32). The generalization of our approach in the presence of a third *observed* system Z, which may be a common driver of X and Y, is presented and discussed in the *SI Appendix*, section 3 and Fig. S1. We do not address in this work the problem of unobserved common drivers, in the presence of which the methodology that we propose carries the risk of incorrectly identifying causal relationships.

Our approach is mainly benchmarked on dynamics which are smooth and deterministic, but it can be applied with no modification to stochastic processes (*SI Appendix*, section 4 and Fig. S2). Since the Information Imbalance is computed over a set of points, we suppose to have access to multiple experiments $X^i(t)$ and $Y^i(t)$ $(i = 1, ..., N)$, representing independent realizations of identical copies of the systems. If the available data consist in a single stationary time series, an ensemble of multiple realizations can be constructed by dividing the trajectory in N nonoverlapping subtrajectories, which can be made efficaciously independent by increasing the time interval between the end of one trajectory and the beginning of the next. In the following, the distance spaces appearing in the Information Imbalance will be labeled by the coordinates employed to construct them. For example, we will use the notation $d_{X(0),Y(\tau)}$ to identify the following Euclidean distance between trajectories i and j:

$$\left(\|X^i(0) - X^j(0)\|^2 + \|Y^i(\tau) - Y^j(\tau)\|^2 \right)^{\frac{1}{2}}.$$

Our causality detection method relies on the intuition that if a dynamic variable X causes another variable Y and one attempts to make a prediction on the future of Y, a distance measure built using the present states of both X and Y will have more predictive

power than a distance built using only Y. This idea is depicted in Fig. 1. Formally, we postulate that if X does not cause Y, then for any $\alpha > 0$

$$\Delta(\alpha) \doteq \Delta\left(d_{\alpha X(0), Y(0)} \to d_{Y(\tau)}\right) > \Delta\left(d_{Y(0)} \to d_{Y(\tau)}\right). \quad [2]$$

Indeed, if Y is autonomous, adding information on the initial value of X can only degrade the information on the future of Y. If, instead, X causes Y, adding information on X will improve the predictability of the future of Y, and $\Delta(\alpha)$ will be minimized by some $\alpha > 0$. The parameter α, as we show in the *SI Appendix, section 5 and Fig. S3*, plays the role of a scaling parameter for the units of X accounting for the magnitude of the coupling strength. τ is a positive parameter representing the time lag of the information transfer from the driving to the driven system. Since this parameter only appears in the argument of the putative driven system Y, the approach is still valid when Y is a dynamical system and X is a static variable (for example a control variable chosen by an experimentalist), with the caveat that in this scenario causality can only be tested in direction $X \to Y$. Using Eq. **2** we can assess the presence of causality by a variational scheme. For this purpose we define the Imbalance Gain in direction $X \to Y$ as

$$\delta\Delta(X \to Y) \doteq \frac{\Delta(\alpha = 0) - \min_\alpha \Delta(\alpha)}{\Delta(\alpha = 0)}. \quad [3]$$

Notice that $\delta\Delta(X \to Y)$ is by construction non-negative. A value of $\delta\Delta(X \to Y) = 0$ indicates that adding the information on the value of X does not help predicting the future of Y, namely Y is autonomous. If instead $\delta\Delta(X \to Y) > 0$, we infer that X causes Y. In this second scenario, we will show that the value of the Imbalance Gain can be used to compare the strengths of different couplings.

Our approach can be viewed as a nonlinear and model-free generalization of Granger causality, as it examines the impact of introducing the supposed causal variable in the past on the predictability of the supposed caused variable in the future.

Ideally, the distance spaces appearing in Eq. **2** should be constructed using all the components of X and Y. However, in real experiments, it is common that not all the variables of each system are recorded. In this case, Takens' theorem (28) ensures that it is possible to recover the information of the missing coordinates by means of the time-delay embeddings of the known variables. For example, if only coordinate $x_1(t)$ is recorded for system X, one can construct the vectors $\widetilde{x}_1(t) = (x_1(t), x_1(t - \tau_e), x_1(t - 2\tau_e), ..., x_1(t - (E-1)\tau_e))$ and the projection of the trajectory in this space is guaranteed to be topologically equivalent to the original orbit for almost any choice of the embedding time τ_e, provided that the embedding dimension E is at least twice larger than the fractal dimension of the original attractor. We highlight that the smallest embedding dimension accomplishing this task is typically smaller in practice: For example, it is well known that the Lorenz attractor can be embedded in a shadow manifold with $E = 3$, while the Takens' theorem would require $E \geq 5$ (33). Even if this mapping is strictly valid only under the assumption of noise-free measurements, it has been empirically demonstrated to be useful also in the analysis of real-world data, which are unavoidably affected by different sources of noise (7, 34). Consistently with our assumptions, the unobserved dynamic variables of each system cannot be unobserved common drivers, as we postulated that neither X nor Y include autonomous subsets of variables. The robustness of our approach with respect to the choice of E, τ_e, and the other relevant hyperparameters (k in Eq. **1** and τ in Eq. **2**) is discussed in the *Materials and Methods*.

Results

We first apply our method to model dynamical systems in which a ground-truth causal relationship is defined, and then we validate
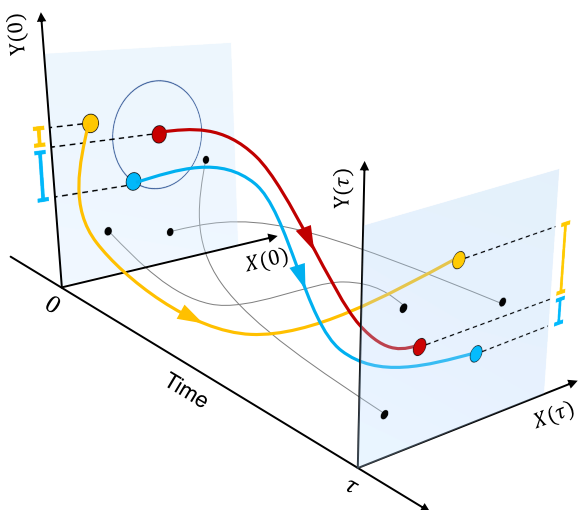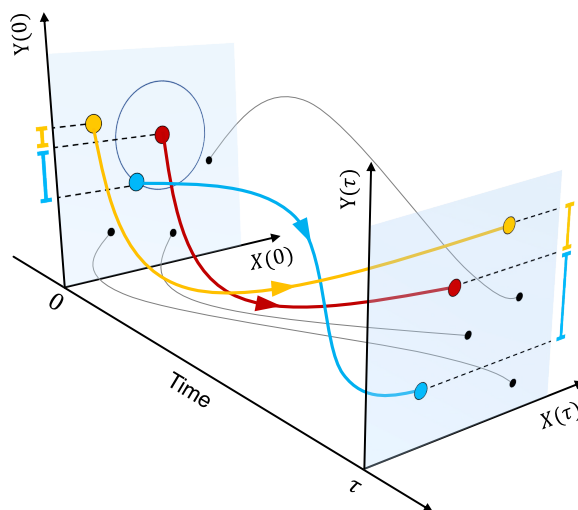

Fig. 1. Illustration of the method applied to the $X \to Y$ direction. Different lines represent independent realizations of a two-dimensional dynamical process. Both in (*A*) and in (*B*) the reference trajectory is depicted in red. At time 0 the blue trajectory is the closest in space $d_{X(0), Y(0)}$, while the yellow realization is the closest in the marginal space $d_{Y(0)}$. When we look at distances in the marginal space $d_{Y(\tau)}$, (*A*) if X causes Y the closest curve is the blue one, which was the closest in the full dynamical space, while (*B*) if X does not cause Y the yellow trajectory remains the closest, as the space $d_{Y(0)}$ already contains the maximal information to predict the state of Y at time τ. Our method tries to predict the k closest trajectories to each realization in the distance space of the putative driven system, assessing the prediction quality as the average of their true distance ranks. The Imbalance Gain in Eq. **3** allows comparing the predictions made in the presence or in absence of the putative causal variables.

it on real-world time series using an EEG dataset collected in our laboratories.

Causality Detection in Model Systems. The dynamical systems employed in the following analysis are based on a first-order dynamics that can be generally written as

$$\dot{X} = f(X), \qquad \text{[4a]}$$

$$\dot{Y} = g(Y) + \varepsilon G(X, Y), \qquad \text{[4b]}$$

in the case of unidirectional coupling $(X \rightarrow Y)$, and

$$\dot{X} = f(X) + \varepsilon_{Y \rightarrow X} F(X, Y), \qquad \text{[5a]}$$

$$\dot{Y} = g(Y) + \varepsilon_{Y \rightarrow X} G(X, Y), \qquad \text{[5b]}$$

in the case of bidirectional coupling $(X \leftrightarrow Y)$. From a conceptual point of view, we highlight that the word "causality" is not used improperly in this context, as these systems satisfy the counterfactual definition that is embraced by modern causal inference in the form of intervention, both in Rubin's potential outcome framework (35) and in Pearl's graphical approach (2). For example, introducing an external forcing in Eq. **4a** would have a clear effect on the dynamics of Y, while disturbing the motion of Y by directly intervening on Eq. **4b** would let the motion of X unperturbed. However, as outlined in the previous Section, our operative approach is based on a predictability principle and does not require the use of external interventions.

In the unidirectional setting, we tested two Rössler systems (36) both with identical and with different frequencies, and two 40-dimensional Lorenz 96 systems (31) with different forcing constants. We tested the bidirectional scenario using two identical Rössler systems and two identical Lorenz systems (37). All the systems display a chaotic dynamics. We refer to the *Materials and Methods* for the explicit equations of each pair of systems. All the tests were performed by extracting from the trajectories $N = 5,000$ realizations, except for the Convergent Cross-Mapping method that requires monitoring the convergence of the results as a function of the number of samples.

In order to benchmark our procedure, we first studied the qualitative behavior of the Information Imbalance $\Delta(\alpha)$ in the optimal scenario where all the coordinates of each system are

known, so that the use of time-delay embeddings is unnecessary. We employed Rössler systems and considered three different coupling configurations, where the link is unidirectional between identical systems (Fig. 2A), unidirectional between different systems (Fig. 2B) and bidirectional between identical systems (Fig. 2C). In these illustrative examples, the number of neighbors k was set to 1 and the time lag τ was fixed to 5. In the case of unidirectional coupling (Fig. 2 A and B), $\Delta(\alpha)$ monotonically increases in the direction where the causal link is absent $(Y \rightarrow X)$, while it clearly shows the presence of a minimum in the correct coupling direction $(X \rightarrow Y)$. Consistently with this scenario, if the coupling is bidirectional the Information Imbalance shows clear minima as a function of α in both directions (Fig. 2C). As a consequence, the Imbalance Gain is positive in both directions.

To demonstrate the robustness of the procedure more quantitatively, in Fig. 3, we report a comparison of our method with four alternative approaches to assess causality between time-dependent variables, namely the Extended Granger Causality (17), Convergent Cross-Mapping (7), the Measure L (29) and Transfer Entropy (23–25), which are described in the (*SI Appendix*, section 2). The latter three approaches are model-free, like ours, while the former assumes a local autoregressive model. Each method produces an estimate for each coupling direction. As the other methods employ time-delay embeddings, to ensure a fair comparison we also applied our approach using the delayed representations of single coordinates (x_1 and y_1). The set of tests was carried out using four different pairs of dynamical systems: two low-dimensional and unidirectionally coupled (identical and different Rössler systems, Fig. 3 A–F and G–L), one low-dimensional and bidirectionally coupled (Lorenz systems, Fig. 3 M–R) and one high-dimensional and unidirectionally coupled (Lorenz 96 systems, Fig. 3 S–X). To evaluate the statistical significance of the results in a consistent way among the different measures, each point in the panels of Fig. 3 was computed as the average of 20 independent estimates, and associated to its SE. The statistical significance of a single Imbalance Gain estimate is assessed with a permutation test over the indices of the putative driver realizations, which generates a null distribution under the hypothesis of absence of causality (*SI Appendix*, section 6 and Table S1).

In the unidirectionally coupled Rössler systems (Fig. 3, first column) and Lorenz systems (Fig. 3, second column) our method
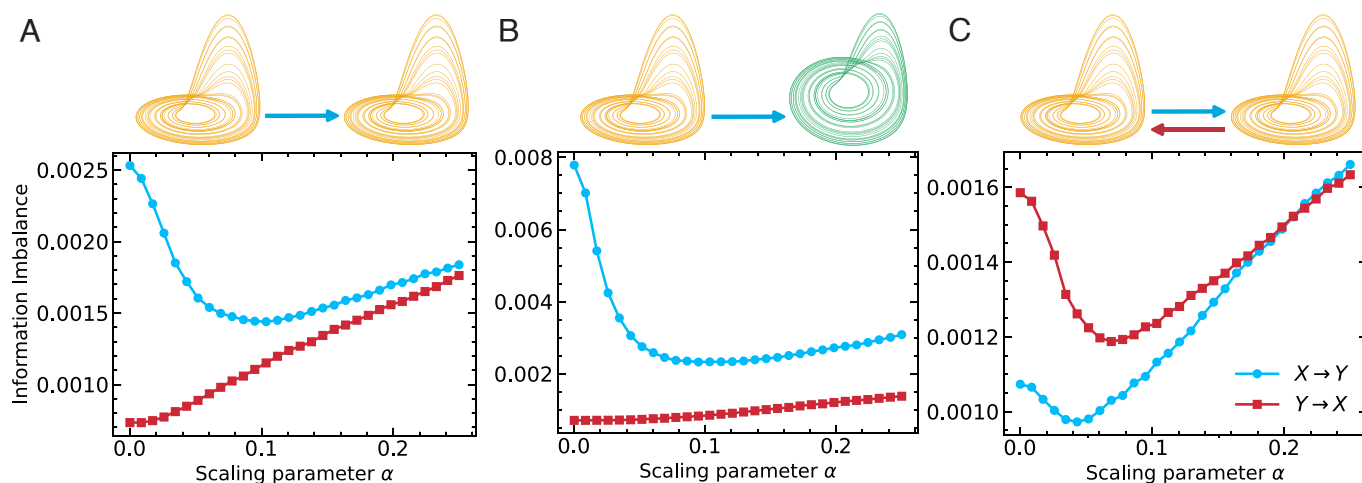

Fig. 2. Profiles of the Information Imbalance $\Delta(\alpha)$ as a function of α, to assess the presence of the causal links $Y \rightarrow X$ and $X \rightarrow Y$. The three panels refer to different pairs of Rössler systems: (A) identical and unidirectionally coupled with coupling strength $\varepsilon = 0.1293$, (B) different and unidirectionally coupled with $\varepsilon = 0.1293$, (C) identical and bidirectionally coupled with $\varepsilon_{X \rightarrow Y} = 0.0603$ and $\varepsilon_{Y \rightarrow X} = 0.1$.
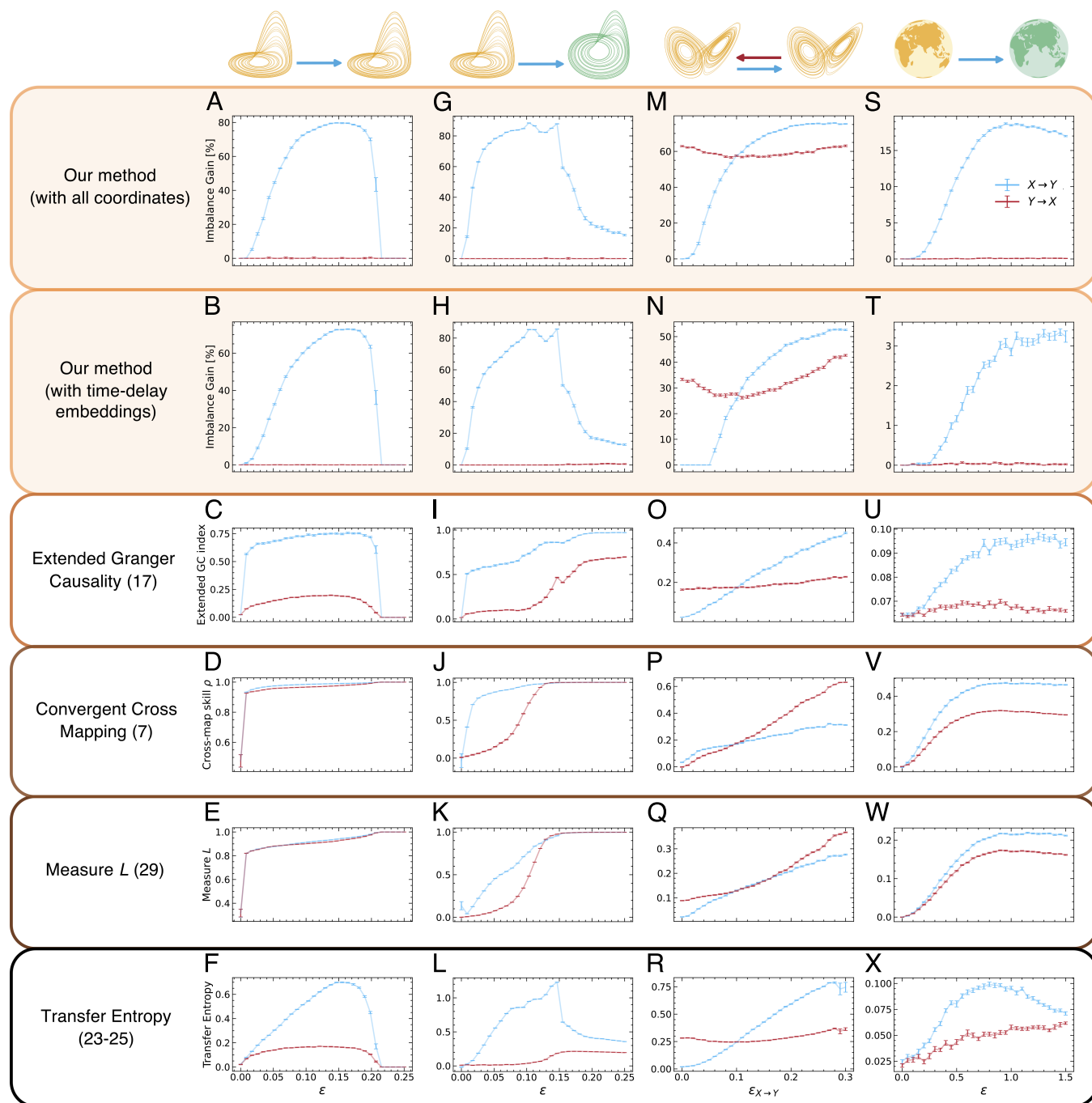
Fig. 3. Comparison of different causality detection methods. The results are shown as a function of the coupling parameter ε ($\varepsilon_{X \to Y}$ in the bidirectional case). (A–F) Identical and unidirectionally coupled Rössler systems. (G–L) Different and unidirectionally coupled Rössler systems. (M–R) Bidirectionally coupled Lorenz systems. The coupling $\varepsilon_{Y \to X}$ was fixed to 0.1. (S–X) Unidirectionally coupled Lorenz 96 systems, with 40 variables each ($F_X = 5$, $F_Y = 6$). In our method, we fixed the embedding time τ_e to 1 and we employed as embedding dimensions (B, H, and N) $E = 3$, and (T) $E = 30$. The time lag was fixed to $\tau = 20$ for the Rössler systems, $\tau = 5$ for the Lorenz systems and $\tau = 30$ for the Lorenz 96 systems.

successfully finds a unidirectional link $X \to Y$, displaying absent or negligible signal in the $Y \to X$ direction. The sharp collapse observed in Fig. 3 A–F at $\varepsilon \sim 0.2$ occur in correspondence of the complete synchronization of the two systems (38, 39), where the trajectories of X and Y become identical. The other methods correctly detect that causality is stronger in the $X \to Y$ direction than in the reverse one; however, they do not allow deducing from the data that the coupling $Y \to X$ is absent.

In the bidirectional case (Fig. 3, third column) all the methods correctly detect the presence of both the causal links, but the cross-mapping methods do not predict the correct ranking of the two coupling strengths for $\varepsilon_{X \to Y} \gtrsim 0.1$. Consistently, in the other methods (Imbalance Gain, Extended Granger Causality and Transfer Entropy) the curves quantifying the strengths of the two causal links intersect at $\varepsilon_{X \to Y} \simeq 0.1$, which is the value at which the opposite coupling parameter $\varepsilon_{Y \to X}$ was fixed. In the high-dimensional scenario (Fig. 3, fourth column), all the approaches detect the correct order of the causal coupling but, once again, the three metrics used for comparison do not allow concluding that causality is actually absent in one direction.

Table 1. False-positive rates (FPR), approximated to the first decimal digit, for detections of couplings $Y \rightarrow X$ in systems coupled in direction $X \rightarrow Y$

	Identical Rössler systems	Different Rössler systems	Lorenz 96 systems
Our method (all coordinates)	0% (0/21)	0% (0/16)	9.7% (3/31)
Our method (embeddings)	0% (0/21)	0% (0/16)	12.9% (4/31)
Extended Granger Causality (17)	100% (21/21)	100% (16/16)	100% (31/31)
Convergent Cross-Mapping (7)	100% (21/21)	93.8% (15/16)	96.8% (30/31)
Measure L (29)	100% (21/21)	93.8% (15/16)	96.8% (30/31)
Transfer Entropy (23–25)	100% (21/21)	100% (16/16)	100% (31/31)

To exclude the results in the synchronization regime, only couplings configurations with $\varepsilon < 0.18$ were considered for the identical Rössler systems, and with $\varepsilon < 0.13$ for the different Rössler systems.

Table 1 reports the number of false-positive detections in the scenarios where the directional coupling $Y \rightarrow X$ is absent, rejecting the null hypothesis of a causal measure being different from zero according to a one-tailed t statistics threshold of $t_{19} = 3.579$ ($P < 0.001$). We report in the *SI Appendix, section 7 and Fig. S4A* an extended description of the statistical test and the false-positive rates for other choices of the significance threshold. The other approaches display a false-positive rate close to 100%. With our measure, the false positives are absent in the Rössler systems, while they are around 10 to 13% in the 40-dimensional Lorenz 96 systems. The abrupt reduction in false-positive detection is a major advantage of our approach.

Causality Detection on EEG Time Series. We employed EEG data to validate our approach in a real-world scenario. We performed a psychophysical experiment to assess whether the Imbalance Gain could establish the presence of a causal relationship between the experimental manipulation and EEG activity across participants, and to understand whether it could be used to study the information flow between different EEG channels. In the experiment 19 healthy volunteers were asked to judge the duration of two stimuli, a visual and an auditory one, presented sequentially. Participants' task was to report, by pressing a key, which of the two stimuli was displayed for longer time. The first stimulus in the pair, the comparison stimulus, was a visual grating varying randomly in display time in each experimental trial. Seven different durations of the visual stimulus were tested, evenly spaced between 0.3 and 0.9 s. The second stimulus, called standard, was a burst of white noise presented through headphones for 0.6 s in each trial. The details of the experiment and additional validation tests are reported in the *SI Appendix, section 8 and Fig. S5*.

First we investigated the causal link between the duration of the comparison stimulus, which is a categorical variable assuming a value between 0.3 and 0.9 s, and the EEG traces relative to its onset and its offset. Notably, in this application the putative causal variable $X(0)$ appearing in Eq. **2** is one-dimensional and time-independent. We excluded from our analysis the information on the second stimulus, as it was shown for a fixed duration among different trials. The Imbalance Gain was computed independently for each participant using the different trials as independent realizations and employing time-delay embeddings of 44 ms ($E = 12$, $\tau_e = 4$ ms). The results are shown for two example channels: one parieto-occipital (POz, Fig. 4A) and one frontal (Fz, Fig. 4B).

We carried out two sets of measurements for different choices of the initial time $t = 0$ appearing in Eq. **2**, which corresponds here to the first point of the predictive delay embedding. In the first tests, we set $t = 0$ to the stimulus onset and we studied the

behavior of the Imbalance Gain $\Delta(\alpha)$ as a function of the time lag τ, limiting τ to a window before the offset of the shortest stimulus. In this setting, trials corresponding to different durations of
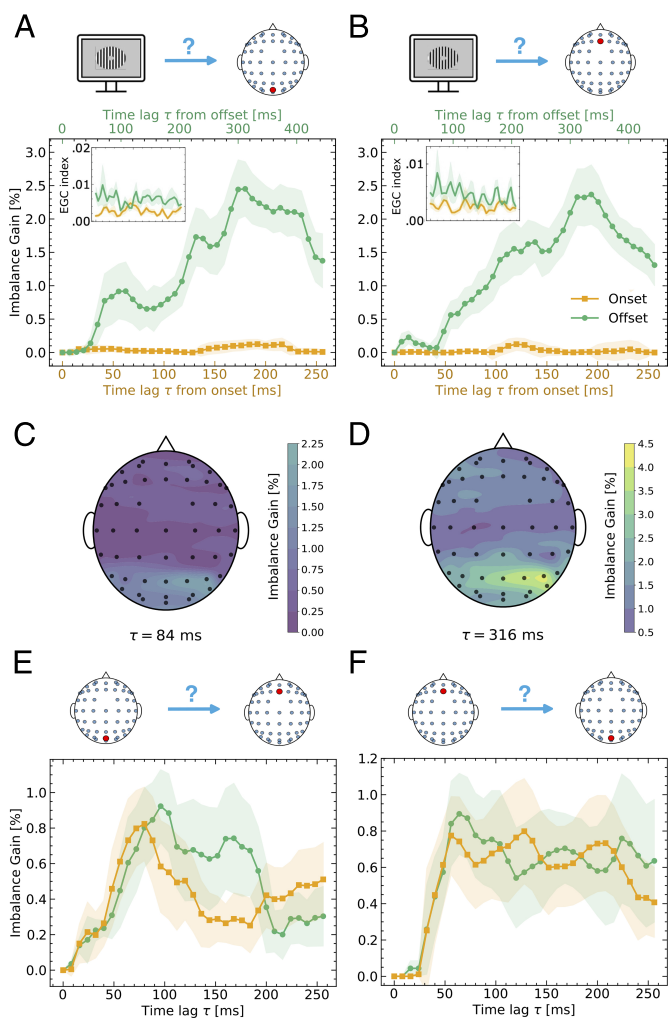

Fig. 4. (*A* and *B*): Imbalance Gains to assess the presence of the link from the duration of the stimulus to the EEG signals in two example channels (*A*) POz and (*B*) Fz, as a function of the time lag τ. The initial time $t = 0$ was set to the onset of the stimulus for the orange curves, and to its offset for the green ones. The shaded area represents the SE associated to the average Imbalance Gain, computed over 19 participants. The inset panels display the average profiles of the Extended Granger Causality index for the same tests and time intervals. (*C* and *D*): Topoplots displaying the distributions of the Imbalance Gains from all the 51 channels, at two different time lags from the stimulus offset. The dynamic topoplots for all the time lags are shown in Movie S1. (*E* and *F*): Imbalance Gains for testing the couplings POz → Fz and Fz → POz, respectively.

the stimulus cannot be yet differentiated because the different comparison stimuli are all indistinguishable in the time window considered; as a consequence, no coupling duration → EEG activity should be observed. Consistently with this observation, we detected only a negligible Imbalance Gain within the first 256 ms from stimulus onset (orange traces in Fig. 4 *A* and *B*). Indeed, using a (one-tailed) t statistics threshold of $t_{18} = 3.610$ to reject the null hypothesis of the Imbalance Gain being equal to zero ($P < 0.001$), we found a total false-positive rate of ∼0.5% across the channels. We report the false-positive rate as a function of the significance threshold in *SI Appendix*, Fig. S4*B*. In the second set of tests, we set the initial time $t = 0$ to the stimulus offset and we studied the Imbalance Gain within a window of 456 ms. From a perceptual stand point, this temporal window represents the period in which a signature of the subjective experience of stimulus duration should arise (40), as only after the offset of the stimulus its duration information becomes available to the participants. Therefore, a possible causal influence between the duration of the stimulus and the EEG activity may emerge in this second scenario. Using the same statistical procedure described above, we detected significant couplings with a different time modulation depending on the channel, whereby the Imbalance Gain started to rise at early latencies in occipital and parietal channels (Fig. 4 *A* and *C*) and peaked in a time span ranging from 300 to 400 ms after the stimulus offset also in frontal channels (Fig. 4 *B* and *D*). This result is in agreement with recent EEG studies in the field of time perception which show that within similar latencies, and particularly in fronto-central electrodes, EEG activity contains information about participants' perceived stimulus duration (41, 42).

As a comparison, we applied the Extended Granger Causality approach to the same causal detection task. We tested different combinations of E and k in order to maximize the difference of the signals in the offset and onset periods, and to minimize at the same time the rate of false positives in the second scenario (see *SI Appendix*, section 2B for details). In the optimal case (*Insets* of Fig. 4 *A* and *B*) we could observe only a faint signal after stimulus offset, and a total false-positive rate of ∼43% after stimulus onset, around 2 orders of magnitude more than with our approach.

Our method is able to retrieve a signature of participants' perceptual decision-making processes and its link to task performance. To illustrate this we studied the causal relationships between two extremes of a hypothetical brain network functionally related to duration processing. Specifically, we evaluated the causal link between a parieto-occipital electrode, POz, and a fronto-central one, Fz. The activity of the former is supposedly linked with an early stage of duration processing (43) where stimulus sensory and duration information is encoded and conveyed to downstream brain regions (duration encoding), while the latter is associated to a higher-level processing stage where duration information is read out and used to perform the task at hand (duration decoding) (44, 45). We computed the Imbalance Gain POz → Fz within 256 ms from stimulus' comparison onset and offset (yellow and green traces in Fig. 4*E*), using time-delay embeddings of 44 ms for both the signals. We found that the Imbalance Gain relative to the different periods (i.e., onset or offset) changed differently as function of τ (period-tau interaction: $F_{26,468} = 1.529$, $P = 0.04$). In both periods the Imbalance Gain peaked around 90 ms, time lag which may reflect the propagation delay in the information flow between the two channels. However, in the onset period, the signal slowly decayed until it reached a plateau, whereas after

stimulus offset we observed a second peak in Imbalance Gain around $\tau = 160$ ms. Interestingly, previous works have shown that, in the same electrode (Fz) and within similar latencies (∼150 ms), it is possible to detect decision-related EEG activity which originates from feed-forward communication from the visual cortices (46, 47). No such effect of the interaction between τ and period on the Imbalance Gain was found in the case Fz → POz ($F_{26,468} = 0.448$, see Fig. 4*F*). To better characterize the relationship between our results and participants' performance we computed the Kendall's correlation coefficient between the Imbalance Gain and participants' accuracy in the task. The results of this analysis show that a positive association is present only at the offset period in correspondence to the Imbalance Gain peaks (in particular around 160 ms, see *SI Appendix*, Fig. S5). Although more experiments are needed to better characterize this result, this shows that the Imbalance Gain is a promising measure to investigate the link between brain functional connectivity and behavior.

Discussion

We proposed an approach to detect causality in time-ordered data based on the Information Imbalance, a statistical measure constructed on distance ranks. The underlying idea is to quantify how the description of a subset of variables in the future is affected by the addition of the putative causal variables in the past; in this sense, our method can be seen as a nonlinear and model-free generalization of Granger causality, suitable for high-dimensional data.

Our approach is also related to the measure L (29), which, using our notation, is based on the comparison of $\Delta\left(d_{X(0)} \to d_{Y(0)}\right)$ with $\Delta\left(d_{Y(0)} \to d_{X(0)}\right)$ (*SI Appendix*, section 2A). This simpler approach faces the limitations illustrated in Fig. 3: the evaluation of a single inequality only allows to identify the dominant causal link, without recognizing situations where the coupling in one direction is absent.

In case of missing dynamic variables, our measure can be applied with time-delay embeddings. In principle, Takens' theorem states that the time-delay embeddings of a single coordinate of Y, e.g., $\widetilde{y}_1$, should allow to reconstruct the full system (X, Y) even when only the coupling $X \to Y$ is present. If this was the case, even when X causes Y there would be no $\alpha > 0$ satisfying Eq. **2** if the distance spaces were constructed with time-delay embeddings, because any information carried by X would be redundant. However, the above statement is correct only if noise is absent, and if it is possible to record the coordinates with arbitrary precision (33). In practical applications, where the observation time is finite and measurement/integration noise is present, the best reconstruction carries only a partial information of X, so that $\widetilde{x}_1$ actually contains unique information about the future of $\widetilde{y}_1$. This scenario is supported by the numerical experiments presented in this work.

By testing our method on several coupled dynamical systems with different coupling configurations, we found that our measure is significantly more robust than the compared methods against the drawback of false positives. In low-dimensional systems the difference in performance with other approaches is striking (Table 1). In the high-dimensional scenario, some false-positive detections are present, but comparing the signals in the two directions allows to clearly discern the irrelevance of one of the two couplings (Fig. 3 *P* and *Q*). The advantage over the other approaches is still evident when the dynamics of driver and

driven systems involve different time scales (*SI Appendix*, section 9 and Fig. S7).

We further applied our measure to real-world time series from EEG experiments, studying the time modulation of causal links from an experimental and static variable—the duration of the comparison stimulus—to the channel-dependent EEG activity, and between EEG activities of different channels. In the tests uncovering causal signals, the time modulations of the Imbalance Gain are consistent with the latencies at which the information of stimulus duration is expected to be transferred from the visual cortex to the fronto-central channels and elaborated by these channels (41, 42). Our findings suggest that applying our approach to ad hoc experiments may provide insights into functional and effective brain connectivity.

We introduced our method within a framework involving two systems X and Y, without considering the potential influence of a third observed system, Z. In the *SI Appendix*, section 3, we show that our approach can be generalized to include such a third system, and how the causal analysis is affected when Z is a common driver ($X \leftarrow Z \rightarrow Y$) or an intermediate system (e.g., $X \rightarrow Z \rightarrow Y$) (48).

In conclusion, we believe that the benchmark presented in this work demonstrates that this approach overcomes relevant limitations of other model-free causality detection methods. Its stronger statistical power shows up in particular when applied to high-dimensional systems in which causality is absent, a scenario which is of utmost importance for understanding how to control a system. This, we believe, will trigger the interest of scientists working on causality detection with real-world time-dependent data.

Materials and Methods

Details on the Dynamical Systems. All the analysis on the dynamical systems reported in *Results* were carried out discarding the first 100,000 samples of the generated time series. All the dynamical systems were integrated using the 8-th order explicit Runge–Kutta method DOP853 in the Python library SciPy, except for the coupled Lorenz 96 systems for which the SciPy implementation of the LSODA integrator was employed (49).

Rössler Systems. In the unidirectional case $X \rightarrow Y$, the equations of the coupled Rössler systems are

$$\dot{x}_1(t) = -\omega_1 x_2(t) - x_3(t), \tag{6a}$$
$$\dot{x}_2(t) = \omega_1 x_1(t) + 0.15 x_2(t), \tag{6b}$$
$$\dot{x}_3(t) = 0.2 + x_3(t) [x_1(t) - 10], \tag{6c}$$
$$\dot{y}_1(t) = -\omega_2 y_2(t) - y_3(t) + \varepsilon (x_1(t) - y_1(t)), \tag{6d}$$
$$\dot{y}_2(t) = \omega_2 y_1(t) + 0.15 y_2(t), \tag{6e}$$
$$\dot{y}_3(t) = 0.2 + y_3(t) [y_1(t) - 10], \tag{6f}$$

with $\omega_1 = \omega_2 = 1.015$ for the case of identical systems, and $\omega_1 = 1.015$, $\omega_2 = 0.985$ for the case of different systems, as studied in refs. 25 and 50. The term $\varepsilon_{Y \rightarrow X} (y_1(t) - x_1(t))$ was added to Eq. **6a** in the bidirectional tests. Both the initial state of system X and the initial state of Y were set by multiplying the components of the vector (10, 10, 10) by three random numbers between 0.5 and 1.5. For all the tested values of the coupling constants, the equations were integrated with a fixed time step of 0.0785 and the time series was downsampled with a frequency of 1/4, resulting in a sampling step of 0.314. After discarding the first 100,000 samples, 105,000 were saved and employed in the analysis.

Lorenz systems. The bidirectionally coupled Lorenz systems are defined by the following equations:

$$\dot{x}_1(t) = 10 (x_2(t) - x_1(t)), \tag{7a}$$
$$\dot{x}_2(t) = x_1(t)(28 - x_3(t)) - x_2(t) + \varepsilon_{Y \rightarrow X} y_1^2(t), \tag{7b}$$
$$\dot{x}_3(t) = x_1(t) x_2(t) - 8/3 x_3(t), \tag{7c}$$
$$\dot{y}_1(t) = 10 (y_2(t) - y_1(t)), \tag{7d}$$
$$\dot{y}_2(t) = y_1(t)(28 - y_3(t)) - y_2(t) + \varepsilon_{X \rightarrow Y} x_1^2(t), \tag{7e}$$
$$\dot{y}_3(t) = y_1(t) y_2(t) - 8/3 y_3(t). \tag{7f}$$

The equations were integrated using the time step $dt = 0.01$, initializing the system with the same protocol used for the Rössler systems. The resulting time series were saved with a downsampling frequency of 1/5, resulting in a sampling step of 0.05 and 205,000 samples.

Lorenz 96 systems. Using the conventions $x_{-1} = x_{N-1}$, $x_0 = x_N$ and $x_{N+1} = x_1$, the equations of the 40-dimensional unidirectionally coupled Lorenz 96 systems employed in the tests are

$$\dot{x}_i = (x_{i+1} - x_{i-2}) x_{i-1} - x_i + F_X, \tag{8a}$$
$$\dot{y}_i = (y_{i+1} - y_{i-2}) y_{i-1} - y_i + F_Y + \varepsilon x_i, \tag{8b}$$

with $i = 1, ..., 40$. The initial state was set to $x_i(0) = F_X (y_i(0) = F_Y)$ for $i > 1$ and to $x_1(0) = F_X + \mathcal{R} (y_1(0) = F_Y + \mathcal{R}')$ for the first component, where $\mathcal{R}$ ($\mathcal{R}'$) is a random number between 0 and 1. The equations were integrated with the time step $dt = 0.03$ and the trajectories were downsampled with frequency 1/2. The tests reported in Fig. 3 were carried out using trajectories with 252,500 points, while the analysis reported in Fig. 5 were carried out over realizations with 752,000 samples.

Average Imbalance Gain. In Figs. 3–5, where we show an average Imbalance Gain, we inferred one optimal scaling parameter α for all the independent estimates. The optimal parameter was obtained by maximizing the α-dependent Imbalance Gain $\langle (\Delta(\alpha = 0) - \Delta(\alpha)) / \Delta(\alpha = 0) \rangle$, where the brackets identify the average across different estimates.

Robustness with Respect to the Hyperparameters. We investigated how the Imbalance Gain is affected by the embedding parameters E and τ_e on a pair of Lorenz 96 systems with $F_X = 6$ and $F_Y = 4$, coupled in direction $X \rightarrow Y$. In Fig. 5, we show the Imbalance Gain in a case in which it should be equal to zero (absence of causality) for different choices of E (along the *x*-axis) and τ_e
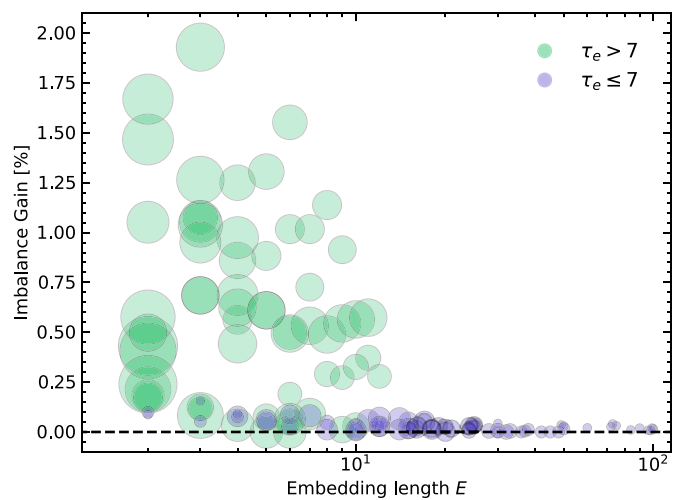

Fig. 5. Imbalance Gain as a function of the embedding length E, computed in direction $Y \rightarrow X$ for two Lorenz 96 systems with opposite coupling $X \rightarrow Y$ ($F_X = 6$, $F_Y = 4$, $\varepsilon = 1$), using $\tau = 30$. The areas of the circles are proportional to the embedding times τ_e employed in the reconstruction. Each point is the average over 20 independent estimates.

(proportional to the area of the circles), restricting to a maximum window length $(E-1)\tau_e = 200$. Due to the absence of a causal link all the points should ideally lie on the dashed line (zero Imbalance Gain). Our measure provides false-positive detections for $E \lesssim 10$, which are particularly evident when the embedding time τ_e is too large ($\tau_e > 7$, green circles). On the other hand, when E is sufficiently large and τ_e is sufficiently small, the Imbalance Gain appears robust against the detection of false positives for several different choices of the embedding parameters ($\tau_e \leq 7$, violet circles). In practical applications, we suggest to fix τ_e to a small value (of the same order of the sampling time) and compute $\delta\Delta$ in a scenario without causality for increasing values of the embedding dimension E, considering the result reliable if $\delta\Delta$ converges to zero. All the analyses shown in this work were performed following this criterion.

Another hyperparameter of our approach is k, the number of neighbors used to compute the Imbalance Gain (Eq. **1**). A large value of k reduces the statistical uncertainty but can bias the estimate toward the absent coupling scenario, as the Imbalance Gain is deterministically equal to 0 in the limit case $k = N - 1$. As a rule of thumb, we suggest a value of k of at most 5 % of the available data. Choosing k in a wide range of values consistent with this choice ($k \sim 10$ to 250 for the Lorenz 96 systems with $N = 5,000$) does not affect the final outcome of the causal analysis in the examples we considered (*SI Appendix*, Fig. S8). All the results in this work were obtained using $k = 1$ for the three-dimensional dynamical systems, and setting $k = 20$ for the Lorenz 96 systems and the EEG analysis.

The last hyperparameter is the value of τ, the delay between the time of observation and the time of the prediction. When the only purpose of the test is assessing the presence or absence of the causal link, the Imbalance Gain provides stable results against different choices of τ (*SI Appendix*, Fig. S9). On the other hand, as shown in the analysis of EEG data, a systematic scan of different values of τ can provide additional insights into the information transfer between driver and driven variables. We underline that the value of τ for which the Imbalance Gain displays a maximum can only be interpreted as the time lag at which a perturbation in the driven system X becomes observable in the driven system Y, and hence the state of the former becomes (maximally) predictive with respect to the state of the latter. This property depends in general both on the actual coupling delay in the interaction $X \to Y$ and on the Lyapunov exponents of the driven system Y. Therefore, the maximum of the Imbalance Gain as a function of τ cannot be taken as a measure of the coupling delay.

As any statistical test, the outcome of our approach depends on the number of samples N, which is a property of the dataset rather than a proper hyperparameter. We show in the *SI Appendix*, Fig. S10 that the scaling of the Imbalance Gain measure as a function of N is comparable to the behaviors of both Measure L and Transfer Entropy, with the advantage of a stable and negligible signal in the direction without coupling in the small N regime.

EEG Experiment. The experiment was conducted on nineteen healthy volunteers (mean age: 24.10, SD: 3.43, 8 females), naïve to the purpose of the experiment, none of them reporting any neurological disease. Participants gave their written informed consent before taking part in the study and were compensated for their participation with 12 Euro/h. The study was carried out in accordance with the Declaration of Helsinki and was approved by the ethics committee of the Scuola Internazionale Superiore di Studi Avanzati (SISSA) protocol number 23970. Further details on the experiment are reported in the *SI Appendix*, section 8.

Statistical assessment on Imbalance Gain data was performed using SciPy (49) and statsmodels (51) packages in Python. In particular, we performed a repeated measures ANOVA to understand the effect of the time lag τ and the period (i.e., onset and offset) on the Imbalance Gain in the case POz $\to$ Fz and Fz $\to$ POz. This analysis was performed using the AnovaRM function with τ and period as within-subjects factors, considering only values of τ larger than 44 ms in order to avoid any overlap between the time-delay embeddings at time 0 and time τ.

1. C. W. J. Granger, Investigating causal relations by econometric models and cross-spectral methods. *Econometrica* **37**, 424–438 (1969).
2. J. Pearl, *Causality: Models, Reasoning and Inference* (Cambridge University Press, ed. 2, 2009).
3. P. Spirtes, K. Zhang, Causal discovery and inference: Concepts and recent methodological advances. *Appl. Inf.* **3**, 1–28 (2016).
4. J. M. Mooij, J. Peters, D. Janzing, J. Zscheischler, B. Schölkopf, Distinguishing cause from effect using observational data: Methods and benchmarks. *J. Mach. Learn. Res.* **17**, 1103–1204 (2016).
5. G. Gendron, M. Witbrock, G. Dobbie, A survey of methods, challenges and perspectives in causality. arXiv [Preprint] (2023). https://doi.org/10.48550/arXiv.2302.00293 (Accessed 14 April 2024).
6. K. D. Hoover, *Causality in Economics and Econometrics* (Palgrave Macmillan UK, London, UK, 2017), pp. 1–13.
7. G. Sugihara et al., Detecting causality in complex ecosystems. *Science* **338**, 496–500 (2012).
8. M. Kretschmer, D. Coumou, J. F. Donges, J. Runge, Using causal effect networks to analyze different arctic drivers of midlatitude winter circulation. *J. Clim.* **29**, 4069–4081 (2016).
9. K. Friston, L. Harrison, W. Penny, Dynamic causal modelling. *NeuroImage* **19**, 1273–1302 (2003).
10. M. Kaminski, M. Ding, W. A. Truccolo, S. L. Bressler, Evaluating causal relations in neural systems: Granger causality, directed transfer function and statistical assessment of significance. *Biol. Cybern.* **85**, 145–157 (2001).
11. S. L. Bressler, A. K. Seth, Wiener–Granger causality: A well established methodology. *NeuroImage* **58**, 323–329 (2011).
12. J. Runge, Causal network reconstruction from time series: From theoretical assumptions to practical estimation. *Chaos* **28**, 075310 (2018).
13. J. Runge et al., Inferring causation from time series in Earth system sciences. *Nat. Commun.* **10**, 2553 (2019).
14. A. Krakovská et al., Comparison of six methods for the detection of causality in a bivariate time series. *Phys. Rev. E* **97**, 042207 (2018).
15. A. E. Yuan, W. Shou, Data-driven causal analysis of observational biological time series. *eLife* **11**, e72518 (2022).
16. E. Beckenbach, N. Wiener, *Modern Mathematics for Engineers* (McGraw-Hill, New York, NY, 1956).
17. Y. Chen, G. Rangarajan, J. Feng, M. Ding, Analyzing multiple nonlinear time series with extended granger causality. *Phys. Lett. A* **324**, 26–35 (2004).
18. D. Marinazzo, M. Pellicoro, S. Stramaglia, Nonlinear parametric model for granger causality of time series. *Phys. Rev. E* **73**, 066216 (2006).
19. D. Marinazzo, M. Pellicoro, S. Stramaglia, Kernel method for nonlinear granger causality. *Phys. Rev. Lett.* **100**, 144103 (2008).
20. L. Barnett, A. K. Seth, Granger causality for state-space models. *Phys. Rev. E* **91**, 040101 (2015).
21. A. Krakovská, F. Hanzely, Testing for causality in reconstructed state spaces by an optimized mixed prediction method. *Phys. Rev. E* **94**, 052203 (2016).
22. M. Wang, Z. Fu, A new method of nonlinear causality detection: Reservoir computing granger causality. *Chaos Solit. Fract.* **154**, 111675 (2022).
23. T. Schreiber, Measuring information transfer. *Phys. Rev. Lett.* **85**, 461–464 (2000).
24. M. Paluš, V. Komárek, Zc. Hrnčíř, K. Štěrbová, Synchronization as adjustment of information rates: Detection from bivariate time series. *Phys. Rev. E* **63**, 046211 (2001).
25. M. Paluš, M. Vejmelka, Directionality of coupling from bivariate time series: How to avoid false causalities and missed connections. *Phys. Rev. E* **75**, 056211 (2007).
26. L. Barnett, A. B. Barrett, A. K. Seth, Granger causality and transfer entropy are equivalent for Gaussian variables. *Phys. Rev. Lett.* **103**, 238701 (2009).
27. J. Runge, V. Petoukhov, J. Kurths, Escaping the curse of dimensionality in estimating multivariate transfer entropy. *Phys. Rev. Lett.* **108**, 258701 (2012).
28. F. Takens, "Detecting strange attractors in turbulence" in *Dynamical Systems and Turbulence, Warwick 1980*, D. Rand, L. S. Young, Eds. (Springer Berlin Heidelberg, Berlin/Heidelberg, Germany, 1981), pp. 366–381.
29. D. Chicharro, R. G. Andrzejak, Reliable detection of directional couplings using rank statistics. *Phys. Rev. E* **80**, 026217 (2009).
30. A. Glielmo, C. Zeni, B. Cheng, G. Csányi, A. Laio, Ranking the information content of distance measures. *PNAS Nexus* **1**, pgac039 (2022).
31. E. N. Lorenz, "Predictability–A problem partly solved" in *Predictability of Weather and Climate*, T. Palmer, R. Hagedorn, Eds. (Cambridge University Press, 2006), pp. 40–58.
32. M. Castro et al., Time series causal relationships discovery through feature importance and ensemble models. *Sci. Rep.* **13**, 11402 (2023).
33. M. B. Kennel, R. Brown, H. D. I. Abarbanel, Determining embedding dimension for phase-space reconstruction using a geometrical construction. *Phys. Rev. A* **45**, 3403–3411 (1992).
34. A. Krakovska, S. Pocos, K. Mojzisova, I. Beckova, J. X. Gubas, State space reconstruction techniques and the accuracy of prediction. *Commun. Nonlinear Sci. Numer. Simul.* **111**, 106422 (2022).
35. G. W. Imbens, D. B. Rubin, *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction* (Cambridge University Press, 2015).
36. O. Rössler, An equation for continuous chaos. *Phys. Lett. A* **57**, 397–398 (1976).
37. E. N. Lorenz, Deterministic nonperiodic flow. *J. Atmos. Sci.* **20**, 130–141 (1963).
38. S. Boccaletti, J. Kurths, G. V. Osipov, D. L. Valladares, C. Zhou, The synchronization of chaotic systems. *Phys. Rep.* **366**, 1–101 (2002).

39. D. Shah, S. Springer, H. Haario, B. Barbiellini, L. Kalachev, Data based quantification of synchronization. *Found. Data Sci.* **5**, 152–176 (2023).

40. T. W. Kononowicz, H. van Rijn, Decoupling interval timing and climbing neural activity: A dissociation between CNV and N1P2 amplitudes. *J. Neurosci.* **34**, 2931–2939 (2014).

41. N. Ofir, A. N. Landau, Neural signatures of evidence accumulation in temporal decisions. *Curr. Biol.* **32**, 4093–4100.e6 (2022).

42. A. Damsma, N. Schlichting, H. van Rijn, Temporal context actively shapes EEG signatures of time perception. *J. Neurosci.* **41**, 4514–4523 (2021).

43. Y. Tonoyan, M. Fornaciai, B. Parsons, D. Bueti, Subjective time is predicted by local and early visual processing. *NeuroImage* **264**, 119707 (2022).

44. F. Protopapa *et al.*, Chronotopic maps in human supplementary motor area. *PLoS Biol.* **17**, 1–34 (2019).

45. F. Protopapa, S. Kulashekhar, M. J. Hayashi, R. Kanai, D. Bueti, Effective connectivity in a duration selective cortico-cerebellar network. *Sci. Rep.* **13**, 1–17 (2023).

46. S. Thorpe, D. Fize, C. Marlot, Speed of processing in the human visual system. *Nature* **381**, 520–522 (1996).

47. M. Fabre-Thorpe, A. Delorme, C. Marlot, S. Thorpe, A limit to the speed of processing in ultra-rapid visual categorization of novel natural scenes. *J. Cogn. Neurosci.* **13**, 171–180 (2001).

48. S. Leng *et al.*, Partial cross mapping eliminates indirect causal influences. *Nat. Commun.* **11**, 2632 (2020).

49. P. Virtanen *et al.*, SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nat. Methods* **17**, 261–272 (2020).

50. M. Paluá, A. Krakovská, J. Jakubík, M. Chvosteková, Causality, dynamical systems and the arrow of time. *Chaos* **28**, 075307 (2018).

51. S. Seabold, J. Perktold, "Statsmodels: Econometric and statistical modeling with Python" in *9th Python in Science Conference* (2010).

52. A. Glielmo *et al.*, DADApy: Distance-based analysis of data-manifolds in Python. *Patterns* **3**, 100589 (2022).