Application of symbolic recurrence to experimental data, from firearm prevalence to fish swimming

Cite as: Chaos **29**, 113128 (2019); doi: 10.1063/1.5119883
Submitted: 14 July 2019 · Accepted: 23 October 2019 ·
Published Online: 26 November 2019

Alain Boldini,[1] Mert Karakaya,[1] Manuel Ruiz Marín,[2] and Maurizio Porfiri[1,3,a)]

AFFILIATIONS

[1]Department of Mechanical and Aerospace Engineering, New York University, Tandon School of Engineering, Brooklyn, New York 11201, USA
[2]Department of Quantitative Methods, Law and Modern Languages, Technical University of Cartagena, 30201 Murcia, Spain
[3]Department of Biomedical Engineering, New York University, Tandon School of Engineering, Brooklyn, New York 11201, USA

[a)]**Author to whom correspondence should be addressed:** mporfiri@nyu.edu

ABSTRACT

Recurrence plots and recurrence quantification analysis are powerful tools to study the behavior of dynamical systems. What we learn through these tools is typically determined by the choice of a distance threshold in the phase space, which introduces arbitrariness in the definition of recurrence. Not only does symbolic recurrence overcome this difficulty, but also it offers a richer representation that book-keeps the recurrent portions of the phase space. Using symbolic recurrences, we can construct recurrence plots, perform quantification analysis, and examine causal links between dynamical systems from their time-series. Although previous efforts have demonstrated the feasibility of such a symbolic framework on synthetic data, the study of real time-series remains elusive. Here, we seek to bridge this gap by systematically examining a wide range of experimental datasets, from firearm prevalence and media coverage in the United States to the effect of sex on the interaction of swimming fish. This work offers a compelling demonstration of the potential of symbolic recurrence in the study of real-world applications across different research fields while providing a computer code for researchers to perform their own time-series explorations.

When investigating the behavior of real-world dynamical systems, the analysis of time-series might be the only practical approach due to the lack of mathematically-tractable models or the high cost of numerical simulations. Symbolic recurrence tools might offer a vantage point for the analysis of time-series by affording intuitive visualization of hidden structures, quantitative measures of important dynamic features, and discovery of causal links between systems. Working with experimental data across a wide range of technical fields, we demonstrate the potential of symbolic recurrence tools in real-world applications. How did gun sales in the United States change over the last 20 years? Were they influenced by the media? What is the role of sex on the interactions between fish? These are some of the questions that we address in this paper toward a systematic assessment of how and when to pursue symbolic recurrence analysis in the study of real-world problems.

I. INTRODUCTION

Recurrence constitutes a powerful lens through which one can analyze the behavior of dynamical systems. While the idea of recurrence dates back to the seminal work of Poincaré,[1] it was not until the effort of Eckmann *et al.*[2] that recurrence plots (RPs) and recurrence quantification analysis (RQA) were introduced. These tools allow researchers to gather valuable information about the behavior of dynamical systems in their phase space, with access only to their time-series.[3,4] RPs and RQA are now part of the methodological toolbox of several fields of investigation, encompassing climatology,[5] geophysics,[6] early damage detection,[7] finance,[8] and neuroscience.[9] Across all these areas of study, the complexity of encountered dynamics hinders the applicability of analytical and numerical techniques.

Despite the demonstrated success of RPs and RQA, they are not free of technical limitations that could ultimately challenge our

ability to draw reliable conclusions regarding the behavior of dynamical systems. Critical to the definition of recurrence is the notion of proximity in the phase space, encapsulated by the so-called proximity parameter ε that discriminates recurrent from nonrecurrent points. Often, the selection of this numerical threshold is not guided by general theoretical principles, but it is executed on a case-by-case basis. Such an arbitrariness leads to uncertainty in the RP and RQA results, possibly begetting misleading or even inaccurate conclusions. Put simply, an excessively large value of ε will overestimate recurrent dynamics, while a small value will underestimate it.

To overcome this difficulty, Caballero-Pintado *et al.*[10] proposed an alternative approach that examines recurrence over a symbolic representation of the time-series. Within this approach, the researcher maps the original time-series to a sequence of discrete symbols that is constructed following ordinal patterns, where each symbol is associated with the local ordering of consecutive values in the time-series. Working with a symbolic representation not only overcomes the arbitrariness of thresholding, but also the black-and-white visualization of classical RPs. In this vein, symbolic RPs (SRPs) offer a colored visualization of recurrence that can help the researcher tease out which portions of the phase space are being visited by the system. While the potential of SRPs has been extensively documented through synthetic datasets in Caballero-Pintado *et al.*,[10] application to experimental data remains elusive, especially in the context of symbolic recurrence quantification analysis (SRQA).

Grounded in the notion of SRPs, Porfiri and Ruiz Marín[11] formulated an information-theoretic backdrop to examine symbolic recurrent dynamics. Building upon previous efforts toward incorporating information-theoretic measures in RPs[12–17] and symbolic representations of time-series,[18,19] the authors established a probability space on which to construct entropy-related measures on symbolic recurrences for univariate, bivariate, and general multivariate processes. Among these measures, they extended the notion of transfer entropy by Schreiber[20] to symbolic recurrences toward guiding the inference of causal influence between dynamical systems from their recurrent behavior. Preliminary results by Porfiri and Ruiz Marín[11] support the feasibility of such an approach to discover causal links, but its full potential is presently untested.

The chief goal of this paper is to fill gaps in knowledge on the use of SRPs, SRQA, and transfer entropy on symbolic recurrences by demonstrating their applicability to real-world time-series, where there is a need to study dynamical features and identify causal links. We begin by illustrating the application of the framework to a sparse dataset related to firearm prevalence in the United States. The dataset comprises two time-series, one for the number of media articles related to firearm laws and regulations and one for the number of background checks at the national level, from January 1999 to December 2017. This dataset[21] serves as a starting point to visualize recurrent dynamics and discover causal links in sparse datasets. Then, we focus on three controlled experiments that encapsulate widely different types of interactions: fluid-structure[22] (fluid-mediated interactions between tandem airfoils), predator-prey[23] (fear response of a live fish exposed to a robotic predator), and human-computer[24] (competition between a human and a virtual opponent in a videogame). Common to these experiments is *a priori* knowledge of existing causal links, which we leverage to test whether transfer entropy on symbolic recurrences can support the discovery of causal influence. We conclude the effort by examining a dataset on animal behavior[25] in which such knowledge is not available toward demonstrating the application of the framework to help answer untapped scientific questions from experimental time-series. We specifically seek to quantify the role of sex on the shoaling behavior of pairs of fish.

The main contributions of this paper include (i) to demonstrate a symbolic framework to study recurrent dynamics associated with firearm prevalence in the United States, from a single pair of short time-series; (ii) to validate the use of SRPs, SRQA, and transfer entropy on symbolic recurrences in the analysis of time-series of controlled experiments, spanning widely different research fields; and (iii) to quantify the role of sex on the social interaction between fish through the study of symbolic recurrent dynamics and the application of transfer entropy on symbolic recurrences. In addition, we present a user-friendly software, which is freely available as the supplementary material to this paper and can be utilized to obtain SRPs, SRQA, and transfer entropy on symbolic recurrences for arbitrary datasets.

The rest of the paper is organized as follows. In Sec. II, we present a concise summary of the methods proposed in Caballero-Pintado *et al.*[10] and Porfiri and Ruiz Marín[11] to investigate the behavior of dynamical systems from their time-series through SRPs, SRQA, and transfer entropy on symbolic recurrences. Section III introduces the dataset on media coverage and background checks in the United States as a primer to illustrate the proposed technical tools on a single pair of time-series where we can tie recurrent features to recent historical events in the United States. In Sec. IV, we explore the validity of the proposed framework through the study of three different datasets, across mechanics and behavioral sciences. In Sec. V, we demonstrate the potential of the framework to tease out the role of sex on social interactions in fish. Finally, Sec. VI outlines the main conclusions of our work and highlights possible future directions. In the Appendix, we describe the functionalities of the provided software.

II. BACKGROUND

Here, we summarize the key technical elements of the framework proposed in Caballero-Pintado *et al.*[10] and Porfiri and Ruiz Marín[11] for the study of dynamical systems through symbolic recurrences and the inference of causal links between them. We begin by introducing the notion of SRPs and tools for SRQA, including measures of determinism and trapping time that are important for describing the structure of one or more dynamical systems. Then, we focus on transfer entropy on symbolic recurrences to support statistically-principled inference of causal links between dynamical systems from their time-series.

A. Symbolic recurrence plots (SRPs)

RPs are a graphical tool to identify the recurrences of the states of a dynamical system in its phase space from time-series. Specifically, given a scalar time-series $\{x_t\}_{t=1}^{T}$ of length T, where T is a positive integer, we can construct its phase space representation by means of Takens's time-delay method[26] for any embedding dimension m, such that the phase space vector at time t is $\bar{x}_t = (x_t, x_{t+1}, \ldots, x_{t+m-1})$. In the phase space, we construct a traditional RP as a Boolean matrix[2]

that encodes proximity of states within an m-dimensional ball of radius ε. For each pair of time instants (t, s), $t, s = 1, 2, \ldots, \bar{T}$, with $\bar{T} = T - m + 1$, we define the entry of the recurrent matrix in the set $\{0, 1\}$ as

$$R^x_{ts} = \begin{cases} 1 & \text{if } \|\bar{x}_t - \bar{x}_s\| < \varepsilon, \\ 0 & \text{otherwise,} \end{cases} \tag{1}$$

where $\|\cdot\|$ is a chosen norm that defines whether two states are neighbors with respect to the proximity parameter ε.

As widely documented,[3,4] traditional RPs are strongly sensitive to the proximity parameter ε. For instance, by choosing a small value of ε, almost none of the phase space vectors would recur, producing an empty RP, thereby hindering the real dynamics of the system. On the other hand, a large value of ε would provide a full recurrence matrix, making the RP ineffective in studying the dynamics of the system. To overcome these drawbacks, Caballero-Pintado *et al.*[10] developed the concept of SRP, which introduces a parameter-free approach to recurrence analysis through symbolic dynamics.

In order to define SRPs, we first introduce a finite partition $\mathcal{P} = \{P_1, P_2, \ldots, P_k\}$ of the m-dimensional phase space with cardinality k, associating each partition set P_i of the phase space with the symbol π^x_i. We naturally define the symbolization mapping S^x from the phase space to the symbol set $\Gamma^x = \{\pi^x_1, \pi^x_2, \ldots, \pi^x_k\}$, such that

$$\bar{x}_t \mapsto S^x(\bar{x}_t) = \pi^x. \tag{2}$$

In this sense, a phase space vector that belongs to the i-th subset of the phase space is mapped to the symbol π^x_i associated with the subset.

Following Caballero-Pintado *et al.*,[10] for each symbol $\tilde{\pi}^x$, we define a corresponding SRP that captures the recurrence in time of that particular symbol, namely,

$$SR^x_{ts}(\tilde{\pi}^x) = \begin{cases} 1 & \text{if } S^x(\bar{x}_t) = S^x(\bar{x}_s) = \tilde{\pi}^x, \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

By book-keeping recurrence on each of the symbols in the symbol set, we offer a richer perspective than Eq. (1), whereby we retain additional knowledge about the location of the phase space that is recurrent. While traditional RPs are black-and-white pictures of the system dynamics, SRPs offer a colored representation in which each color tells about the recurrence of a set in the partition of the phase space. Notably, this approach shares similarities with heterogeneous recurrence monitoring,[27] which employs a fractal dimension representation to distinguish among recurrent states.

Should one be interested in obtaining a symbolic black-and-white representation, similar to a traditional RP, they could consolidate recurrence on the entire symbol set within the matrix

$$SR^x = \sum_{i=1}^{k} SR^x(\pi^x_i). \tag{4}$$

This matrix provides an overall measure of recurrence, resembling previous work in the literature on symbolic recurrences.[28–30]

Similar to Caballero-Pintado *et al.*[10] and Porfiri and Ruiz Marín,[11] we focus on a symbolization map based on ordinal patterns. Specifically, for any $m > 1$, we define $k = m!$ symbols associated with the relative order of the entries in the phase space vector $\bar{x}_t$. Such a symbol set is isomorphic to the symmetric group S_m, generated by all possible permutations of m elements. In particular, we denote each symbol with $\pi^x := (i_1, i_2, \ldots, i_m)$, where the entries are distinct natural numbers from 0 to $m - 1$. A phase space vector $\bar{x}_t$ is mapped to this symbol if

$$x_{t+i_1} \leq x_{t+i_2} \leq \cdots \leq x_{t+i_m} \tag{5a}$$

and

$$i_{s-1} < i_s \quad \text{if } x_{t+i_{s-1}} = x_{t+i_s}. \tag{5b}$$

These two conditions ensure ordinal patterning and uniqueness of the symbolization map, respectively.

Working with a symbolic representation, recurrence implies that phase space vectors share the same ordinal pattern, without necessarily being proximal in the sense of classical RPs. This is expected to mitigate the effect of noise, whereby moderate noise levels would not be sufficient to break the order of the components of the phase space vectors, especially when dealing with low embedding dimensions. Long ordinal patterns may be more prone to noise, albeit their practical use is typically challenged by the excessive computational times and the need for unrealistically long time-series.

When considering two dynamical systems with phase space vectors $\bar{x}_t$ and $\bar{y}_t$, we can define the SRP over the joint $\bar{z}_t = (\bar{x}_t, \bar{y}_t)$, where the symbolization is performed component-wise. The SRP over the joint is related to the SRPs for the single components through

$$SR^z_{ts}(\tilde{\pi}^x, \tilde{\pi}^y) = SR^x_{ts}(\tilde{\pi}^x) SR^y_{ts}(\tilde{\pi}^y). \tag{6}$$

An equivalent joint representation can be carried out for more than two dynamical systems.

Although SRPs address the arbitrariness of thresholding in RPs and offer a colored visualization of recurrence, pursuing a symbolic approach does not come without limitations. Specifically, a form of arbitrariness is introduced by the choice of the partition of the phase space and the consequent symbolization, whereby in principle different partitions may produce different results. In addition, the symbolized time-series encapsulates less information with respect to the original, continuous-range time-series. Intuitively, the use of a symbolic approach should be recommended when dealing with time-series that contain easily identifiable patterns and clusters, but these tools may also be applied to more complex dynamical systems.

B. Symbolic recurrence quantification analysis (SRQA)

In order to highlight features of the examined dynamical systems and compare their behavior, SRQA introduces a set of quantitative measures over SRPs,[10] based on indicators widely used in the RPs' community.[3,4] First of all, we define the symbolic recurrence rate (*SRR*) to a given symbol from the SRPs in Eq. (3). For a given symbol $\tilde{\pi}^x$, *SRR* measures the probability of its recurrence, such that

$$SRR(\tilde{\pi}^x) = \frac{1}{\bar{T}(\bar{T} - 1)} \sum_{\substack{t,s=1 \\ t \neq s}}^{\bar{T}} SR^x_{ts}(\tilde{\pi}^x), \tag{7}$$

where we have excluded the diagonal of $SR(\tilde{\pi}^x)$ from the computation. Similar to Eq. (4), it is possible to define an overall *SRR*, which

is related to the *SRR* of all the given symbols by

$$SRR = \sum_{i=1}^{k} SRR(\pi_i^x). \tag{8}$$

This indicator measures the probability of recurrence to any symbol. As a reference, the expected value of the SRR for an independent identically distributed time-series[31] is $1/m!$.

All the other measures considered herein are defined for the aggregated SRP in Eq. (4). These measures are based on two kinds of structures in the SRP, namely, diagonal and vertical (or equivalently horizontal, for symmetry) lines. Specifically, we define the distributions $\{(d, n(d))\}$ and $\{(v, n(v))\}$ of diagonal and vertical lines, where $n(d)$ and $n(v)$ are the number of diagonal and vertical lines of length d and v, respectively. In both cases, we neglect isolated points by setting $d_{\min} = v_{\min} = 2$. Also, when examining the diagonal lines of the matrix, we neglect the main diagonal that is always composed of all ones.

Structures in SRPs provide useful indications about the dynamics of the system. Diagonal lines of length d between (t, s) and $(t + d, s + d)$ identify a sequence of symbols whose phase space vectors $(\bar{x}_t, \bar{x}_s), (\bar{x}_{t+1}, \bar{x}_{s+1}), \ldots, (\bar{x}_{t+d}, \bar{x}_{s+d})$ belong to the same set in the partition of the phase space. Such a set could potentially vary along the sequence, whereby the two phase vectors could visit multiple subsets of the partition. In other words, diagonal lines represent two equal sequences of symbols of length d, one starting at time t and one at time s. Therefore, diagonal lines reveal some form of predictability of the system, with long diagonals indicating long repeated symbols' sequences. We utilize the maximum diagonal length $d_{\max}$ as a measure of recurrence of symbols' sequences, and we use the average diagonal length $\bar{d}$ to elucidate the mean length of the repeated sequences of symbols. From the distribution of the diagonal lines, we can compute the probability that any point belongs to a diagonal line,

$$D = \frac{1}{\bar{T}(\bar{T}-1)} \sum_{i=d_{\min}}^{d_{\max}} d_i n(d_i), \tag{9}$$

where, once again, we have neglected the main diagonal. This indicator is equivalent to the so-called determinism, scaled by the *SRR*.

Another important type of structure observed in SRPs is vertical lines, which correspond to a sequence of phase space vectors $\bar{x}_t, \bar{x}_{t+1}, \ldots, \bar{x}_{t+v}$ belonging to the same set of the partition of the phase space. Similar to classical RPs, the length of the vertical lines quantifies the number of time instants in the time-series where the phase space vectors are constrained in the same set of the symbolic partition. We define the trapping time as the average vertical length $\bar{v}$.

While these measures have been introduced with respect to one dynamical system, they could be readily extended to two or more systems by simply looking at a joint SRP. In this sense, the length of the diagonal lines is associated with the determinism of the joint dynamics of the systems, while the vertical lines measure the trapping time of the joint dynamics.

C. Transfer entropy on symbolic recurrences

The entropy $H(X)$ of a discrete random variable X measures the amount of uncertainty on its outcome.[32] If $\mathcal{X}$ is the sample space of X, then

$$H(X) := -\sum_{x \in \mathcal{X}} \Pr(X = x) \log \Pr(X = x), \tag{10}$$

where we have used the natural logarithm, so that entropy is measured in "nats."

This definition can be extended to the joint and conditional entropies for two random variables X and Y by observing that Eq. (10) is associated with the expectation of the logarithm of the probability distribution, such that[32]

$$H(X, Y) := -\sum_{x \in \mathcal{X}, y \in \mathcal{Y}} \Pr(X = x, Y = y) \log \Pr(X = x, Y = y), \tag{11a}$$

$$H(X|Y) := -\sum_{x \in \mathcal{X}, y \in \mathcal{Y}} \Pr(X = x, Y = y) \log \Pr(X = x|Y = y), \tag{11b}$$

where $\mathcal{Y}$ is the sample space for Y. Joint and conditional entropies indicate the entropy associated with the random variable (X, Y) and the amount of uncertainty in X that cannot be explained by Y, respectively, and are related through

$$H(X|Y) = H(X, Y) - H(Y). \tag{12}$$

By extending these notions to stochastic processes, we can introduce the information-theoretic construct of transfer entropy between two dynamical systems. Specifically, given two stationary processes X_t (target) and Y_t (source), transfer entropy $\mathrm{TE}^{y \to x}$ measures the reduction in the uncertainty of the future state of the target, X_{t+1}, from its present state, X_t, due to additional knowledge regarding the present state of the source, Y_t. For processes with finite sample space, this non-negative measure is defined as[20]

$$\begin{aligned} \mathrm{TE}^{y \to x} &:= H(X_{t+1}|X_t) - H(X_{t+1}|X_t, Y_t) \\ &= \sum_{\substack{x_+ \in \mathcal{X} \\ x \in \mathcal{X} \\ y \in \mathcal{Y}}} \Big[\Pr(X_{t+1} = x_+, X_t = x, Y_t = y) \\ &\quad \times \log \frac{\Pr(X_{t+1} = x_+|X_t = x, Y_t = y)}{\Pr(X_{t+1} = x_+|X_t = x)} \Big]. \end{aligned} \tag{13}$$

Transfer entropy can be computed from the time-series $\{x\}_{t=1}^{T}$ and $\{y\}_{t=1}^{T}$ by means of plug-in estimators of the probability distribution for the generic triplet, that is $(X_{t+1} = x_+, X_t = x, Y_t = y)$.

To calculate transfer entropy on symbolic recurrences, the appropriate probability space has to be established for stochastic processes in the phase space. Given the time-series $\bar{X}_t$ in the phase space, we define the probability that a symbol recurs after at least τ time steps as

$$P_{\mathrm{Rec}}^x(\tilde{\pi}^x, \tau) := \Pr(S^x(\bar{X}_t) = S^x(\bar{X}_s) = \tilde{\pi}^x, |t - s| > \tau). \tag{14}$$

Notably, P_{Rec}^x does not establish a probability distribution over the symbolic space, whereby a symbol may not recur after at least τ time steps with a finite probability. To overcome this problem, we condition Eq. (14) on the recurrence to any symbol in the symbolic space,

obtaining

$$P^{x\star}_{\mathrm{Rec}}(\tilde{\pi}^x,\tau) := P^{x}_{\mathrm{Rec}}(\tilde{\pi}^x|\Gamma^x,\tau) = \frac{P^{x}_{\mathrm{Rec}}(\tilde{\pi}^x,\tau)}{\sum_{\delta_x\in\Gamma^x} P^{x}_{\mathrm{Rec}}(\delta_x,\tau)}. \tag{15}$$

The generalization to bivariate and multivariate cases requires some care in accounting for partial recurrence of only one of the processes. Considering for simplicity the bivariate case with two stochastic processes $\bar{X}_t$ and $\bar{Y}_t$ in the phase space, from Eq. (14) we define the probability that $\bar{X}_t$ and $\bar{Y}_t$ recur to symbols $\tilde{\pi}^x$ and $\tilde{\pi}^y$ after at least τ time steps as

$$P^{xy}_{\mathrm{Rec}}(\tilde{\pi}^x,\tilde{\pi}^y,\tau) := \Pr(S^x(\bar{X}_t) = S^x(\bar{X}_s) = \tilde{\pi}^x, S^y(\bar{Y}_t) = S^y(\bar{Y}_s) = \tilde{\pi}^y, |t-s| > \tau). \tag{16}$$

However, a marginalization of Eq. (16) with respect to Γ^y does not provide the same probability distribution as in Eq. (14), whereby the marginalization assumes the recurrence of symbol $S^y(\bar{Y}_t)$. In other words, the marginalization neglects terms in which only symbol $\tilde{\pi}^x$ recurs after at least τ time steps, that is,

$$P^{xy}_{\mathrm{Rec}}(\tilde{\pi}^x,0,\tau) := \Pr(S^x(\bar{X}_t) = S^x(\bar{X}_s) = \tilde{\pi}^x, S^y(\bar{Y}_t) \neq S^y(\bar{Y}_s), |t-s| > \tau). \tag{17}$$

From this observation, we can calculate a distribution over the symbol space $\Gamma^x \times \Gamma^y$ by conditioning over the recurrence of both processes to a symbol such that

$$P^{xy\star}_{\mathrm{Rec}}(\tilde{\pi}^x,\tilde{\pi}^y,\tau) := P^{xy}_{\mathrm{Rec}}((\tilde{\pi}^x,\tilde{\pi}^y)|\Gamma^x\times\Gamma^y,\tau) = \frac{P^{xy}_{\mathrm{Rec}}(\tilde{\pi}^x,\tilde{\pi}^y,\tau)}{\sum_{\substack{\delta^x\in\Gamma^x\\ \delta^y\in\Gamma^y}} P^{xy}_{\mathrm{Rec}}(\delta^x,\delta^y,\tau)}. \tag{18}$$

The extension to the multivariate case naturally follows from the bivariate case.

Based on these definitions, Porfiri and Ruiz Marín[11] introduced the following definition of transfer entropy on symbolic recurrences:

$$\mathrm{TE}^{y\to x}_{\mathrm{Rec}}(\tau) = \sum_{\substack{\delta^x_+\in\Gamma^x\\ \delta^x\in\Gamma^x\\ \delta^y\in\Gamma^y}} \left[P^{x+xy\star}_{\mathrm{Rec}}(\delta^x_+,\delta^x,\delta^y,\tau) \log \frac{P^{x+xy\star}_{\mathrm{Rec}}(\delta^x_+|(\delta^x,\delta^y),\tau)}{P^{x+xy\star}_{\mathrm{Rec}}(\delta^x_+|\delta^x,\tau)} \right], \tag{19}$$

where

$$P^{x+xy\star}_{\mathrm{Rec}}(\tilde{\pi}^x_+|(\tilde{\pi}^x,\tilde{\pi}^y),\tau) = \frac{P^{x+xy\star}_{\mathrm{Rec}}(\tilde{\pi}^x_+,\tilde{\pi}^x,\tilde{\pi}^y,\tau)}{\sum_{\delta^x_+\in\Gamma^x} P^{x+xy\star}_{\mathrm{Rec}}(\delta^x_+,\tilde{\pi}^x,\tilde{\pi}^y,\tau)}, \tag{20a}$$

$$P^{x+xy\star}_{\mathrm{Rec}}(\tilde{\pi}^x_+|\tilde{\pi}^x,\tau) = \frac{\sum_{\delta^y\in\mathcal{Y}} P^{x+xy\star}_{\mathrm{Rec}}(\tilde{\pi}^x_+,\tilde{\pi}^x,\delta^y,\tau)}{\sum_{\substack{\delta^x_+\in\Gamma^x\\ \delta^y\in\Gamma^y}} P^{x+xy\star}_{\mathrm{Rec}}(\delta^x_+,\tilde{\pi}^x,\delta^y,\tau)}. \tag{20b}$$

Throughout this paper, we consider the asymptotic approximation $\tau \to \infty$. This assumption reduces the computational time, such that any two time steps can be treated as independent. With reference to the time-series $\bar{X}_t$, we estimate the probability of recurrence to symbol $\tilde{\pi}^x$ as $\frac{1}{T^2}\sum_{t,s=1}^{T} SR^x_{ts}(\tilde{\pi}^x)$, which is equivalent to the symbolic recurrence rate in Eq. (7) for T sufficiently large. When considering multiple time-series, one performs an equivalent estimation in terms of the product of SRPs like in Eq. (6).

D. Statistical tests on transfer entropy

Transfer entropy on symbolic recurrences can be utilized to infer causality in a Wiener-Granger sense between two dynamical systems X_t and Y_t. Specifically, the following null and alternative hypotheses can be tested using statistical methods to quantify influence and its direction:

($H1$) $\mathrm{TE}^{x\to y}_{\mathrm{Rec}}(\tau) = 0$ against $\mathrm{TE}^{x\to y}_{\mathrm{Rec}}(\tau) > 0$;

($H2$) $\mathrm{TE}^{y\to x}_{\mathrm{Rec}}(\tau) = 0$ against $\mathrm{TE}^{y\to x}_{\mathrm{Rec}}(\tau) > 0$; and

($H3$) $\mathrm{TE}^{x\to y}_{\mathrm{Rec}}(\tau) - \mathrm{TE}^{y\to x}_{\mathrm{Rec}}(\tau) = 0$ against $\mathrm{TE}^{x\to y}_{\mathrm{Rec}}(\tau) - \mathrm{TE}^{y\to x}_{\mathrm{Rec}}(\tau) \neq 0$.

If ($H1$) is rejected, a causal link from X_t to Y_t is discovered. Similarly, if ($H2$) is rejected, a causal link from Y_t to X_t is discovered. If ($H3$) is rejected, one can infer the main direction of influence between the two systems, where a positive net transfer entropy suggests that X_t influences Y_t more than Y_t influences X_t. Vice versa, if net transfer entropy is negative, Y_t influences X_t more than X_t influences Y_t. Put simply, ($H1$) and ($H2$) can be used to identify the existence of information transfer and ($H3$) to determine its dominant direction. In order to test these hypotheses, we generate surrogate data, which allow us to create a baseline that discriminates true information transfer from noise. Depending on the number of available realizations, we select different methods to generate surrogate data and to test the hypotheses.

In the case of a single pair of time-series, surrogate data can be created through bootstrap methods, as detailed in Porfiri and Ruiz Marín.[11] In this paper, we use the Iterative Amplitude Adjusted Fourier Transform (IAAFT) algorithm[33] to generate multiple time-series from the original one, while preserving its linear structure. After this procedure, the 20 000 pairs of surrogate time-series are symbolized, and transfer entropy on symbolic recurrences for surrogate data is computed as in Sec. II C. We compare the transfer entropy from the original time-series against the empirical distribution of surrogate data using a 5% significance level for both one-tailed ($H1$ and $H2$) and two-tailed ($H3$) tests.

When dealing with many realizations, surrogate data can be generated through permutation tests.[34–36] Given the realizations of two stochastic processes X_t and Y_t, we randomly pair all the realizations of X_t with the ones of Y_t. We compute the mean value of transfer entropy on symbolic recurrences from X_t to Y_t, from Y_t to X_t, and of their difference over the shuffled pairs. Empirical null distributions for the mean of the transfer entropies are obtained by repeating this procedure 20 000 times. In this case, mean transfer entropy values from the real datasets are compared to the empirical distributions of mean transfer entropy values from surrogate data. We perform one-tailed tests for ($H1$) and ($H2$) and two-tailed tests for ($H3$), using in both cases a 5% significance level.

III. ILLUSTRATION OF THE FRAMEWORK ON FIREARM PREVALENCE DATA

To illustrate our strategy to SRPs, SRQA, and transfer entropy on symbolic recurrences summarized in Sec. II, we consider a pair of time-series from Porfiri *et al.*,[21] representing the number of documents on firearm laws and regulations published in The New York Times and The Washington Post, and the number of background

checks to acquire new firearms in the United States, every month from January 1999 to December 2017.

The number of documents published in these venues can be regarded as a measure of the societal debate regarding gun control policies in the United States. Through the media, people learn about potential legislative changes that may be ahead of them, potentially challenging their ability to acquire firearms in the future. The number of background checks is an indirect measure of the tendency of people to purchase firearms, whereby most of firearm acquisition automatically triggers a background check at the Federal level.

The traditional transfer entropy analysis by Porfiri *et al.*[21] demonstrated that media output has a significant influence over background checks, supporting that people could fear that their rights to purchase firearms could be challenged. Interestingly, the traditional transfer entropy analysis by Porfiri *et al.*[21] did not point at an influence of background checks on media output, suggesting that other explanatory variables determine media coverage on firearm regulation, potentially including gruesome firearms violence events.

In the present analysis, we denote the two datasets as "media output" and "background checks." Consistent with Porfiri *et al.*,[21] we work with detrended and seasonally-adjusted data for background checks, which would be otherwise dominated by the typical cycle of firearm acquisition at the end of the year and in the spring, along with the steady increase we have witnessed over the last 20 years. Given the lack of multiple realizations, the IAAFT algorithm is used to generate surrogate data.

For clarity, we only visualize SRPs for time-series of 150 months, commencing on January 2003, but SRQA and transfer entropy computations are performed over the entire time-series of 228 time steps. Figure 1 shows the original and symbolic time-series for media output and background checks for an embedding dimension $m = 2$, such that a decrease (increase) in a variable from a month to the next one is associated with symbol π_1 (π_2) and indicated in the plot with a blue (red) dot. With 228 points in the time-series, we cannot utilize values of m above 2 in the transfer entropy computation. In fact, going up to $m = 3$ would require estimating $(m!)^3 = 216$ values of the probability mass function, which would be difficult to do with only 228 points. Figure 2 illustrates the SRPs for media output, background checks, and their joint for the same embedding dimension. From Fig. 2(a), we identify a prominent blue vertical (horizontal) band between January and September 2013, which correspond to the aftermath of the Sandy Hook mass shooting, where 27 people, including 20 children, lost their lives. The massacre triggered a profound debate on firearm control, with more than 1000 articles written within the month of January 2013. In the next few months, the media output decreased, which is seen by the blue band associated with the symbol π_1^x (decreasing ordinal pattern). The same band is evident in the SRP of the background checks in Fig. 2(b) and in the

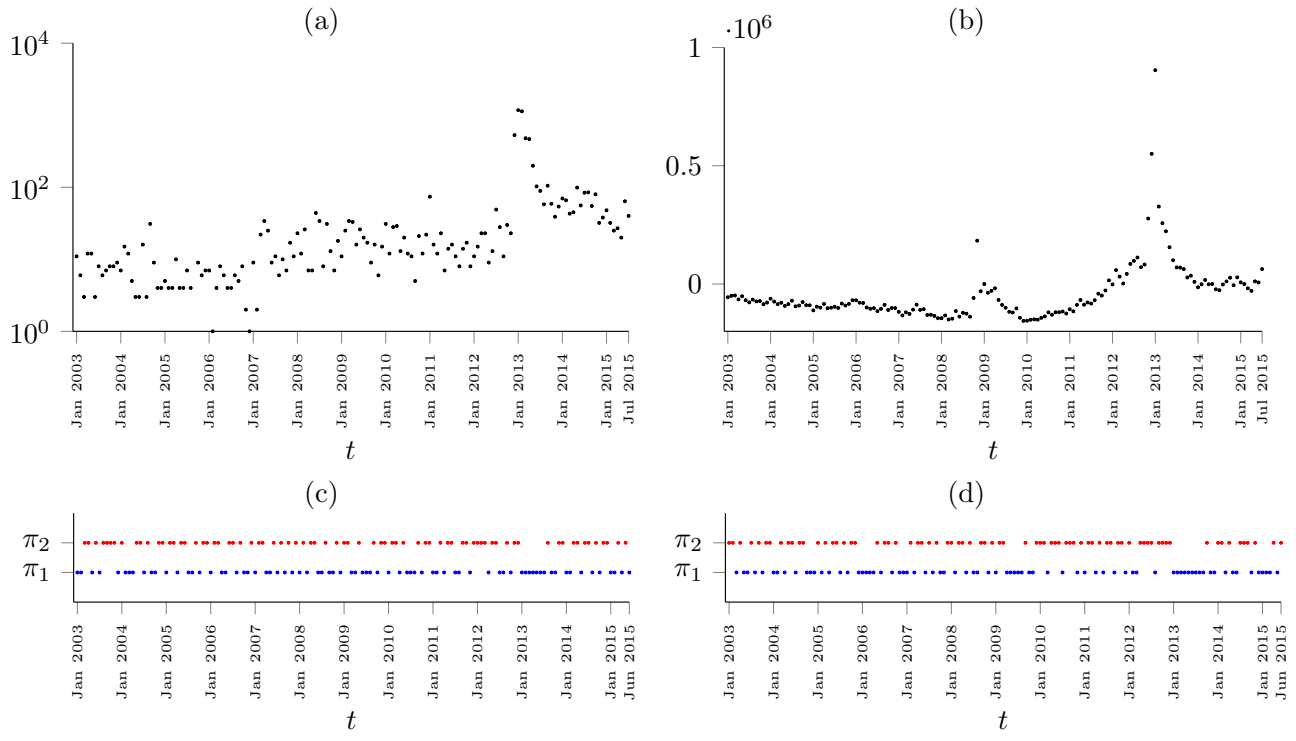

FIG. 1. Time-series of (a) media output and (b) background checks, along with their corresponding symbolic time-series (c) and (d), respectively, for an embedding dimension of $m = 2$. π_1 and π_2 indicates decreasing and increasing ordinal patterns, respectively. Note that, for $m = 2$, symbolic time-series have one point less than the original time-series.

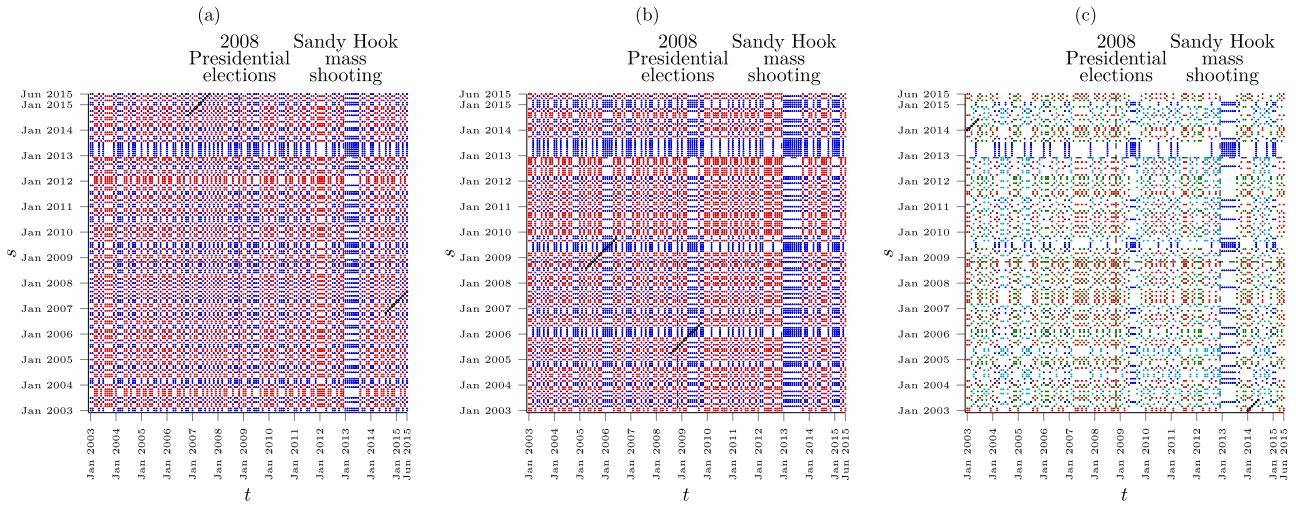

FIG. 2. SRPs of (a) media output, (b) background checks (seasonally-adjusted and detrended), and (c) their joint distribution. In (a) and (b), blue and red dots indicate decreasing (π_1) and increasing (π_2) ordinal patterns, respectively. In (c), blue, red, dark green, and cyan indicate decreasing-decreasing (π_1^x, π_1^y), decreasing-increasing (π_1^x, π_2^y), increasing-decreasing (π_2^x, π_1^y), and increasing-increasing (π_2^x, π_2^y) ordinal patterns for the joint symbols, respectively. Black dots mark the longest diagonal line in each SRP.

joint SRP in Fig. 2(c), supporting the link between the people tendency to buy firearms and media output on regulation proposed by Porfiri *et al.*[21]

Interestingly, another blue band can be identified in the SRP of background checks, corresponding to the five-month period from April 2009 to August 2009 after the election of President Obama in November 2008 and the Carthage nursing home shooting in March 2009. This period follows another period of several months that saw a rise in firearm acquisition, as empirically demonstrated by Depetris-Chauvin,[37] possibly due to fear of stricter laws and regulations on firearms.

Increments in both media output and background checks, corresponding to the red dots in Fig. 2, present a much more erratic behavior. This may be explained by the nature of media coverage, which might rapidly increase in the aftermath of some tragic events and slowly decay over time, explaining the different structures for the two symbols (π_1^x and π_2^x). Background checks follow a similar pattern, as illustrated in the joint SRP in Fig. 2(c), with some exceptions possibly related to long term political trends on firearm regulations. In this sense, a remarkable structure is seen between April and December 2012, a period that registered a steady increase in background checks, with only August breaking this trend. This pattern is possibly related to the incumbent reelection of President Obama in November 2012.

The SRQA of this dataset is summarized in Table I (for simplicity, in the tables, we omit time when referring to the processes). For both media output and background checks, the symbol associated with a decrease between two consecutive months recurs slightly more than the one marking an increase. Overall, the symbolic recurrence rates of the symbols sum up to around 0.5, similar to an independent identically distributed time-series. However, the two SRPs present some form of predictability, whereby the longest diagonal lines span more than one year. On the other hand, the mean lengths of diagonal lines have a modest value, suggesting that long repeated sequences of symbols pertain to limited events. While the system remains predictable for short periods of time, almost 40% of the points in the SRPs (including white dots) belong to diagonal lines. The trapping time is slightly higher for background checks than for media output, confirming our intuition that background checks may follow longer general political climate on firearm regulations.

TABLE I. SRQA of the dataset on firearm prevalence in the United States, illustrating the measures defined in Sec. II B. Here, *X* indicates media output, *Y* refers to background checks, and (*X*, *Y*) is their joint process.

	$SRR(\pi_1)$			$SRR(\pi_2)$		$d_{\max}$	$\bar{d}$	D	$\bar{v}$
X	0.260			0.238		17	3.101	0.384	2.397
Y	0.274			0.225		15	3.057	0.377	2.679
	$SRR(\pi_1^x, \pi_1^y)$	$SRR(\pi_1^x, \pi_2^y)$	$SRR(\pi_2^x, \pi_1^y)$	$SRR(\pi_2^x, \pi_2^y)$		$d_{\max}$	$\bar{d}$	D	$\bar{v}$
(X, Y)	0.067	0.062	0.069	0.050		9	2.380	0.118	2.357

The joint SRP shows almost equal recurrence rates among the four symbols, with an overall *SRR* covering around 25% of the points in the SRP. The maximum diagonal length, the average diagonal length, and the determinism have lower values when compared to the SRPs of the single variables, suggesting that the joint variable is less predictable than the single ones. The trapping time, instead, is similar to the trapping time of media output, as one might expect from the higher variability of media output with respect to background checks.

The results of transfer entropy analyses are shown in Fig. 3 (in the figures about transfer entropy, we use a simpler notation than Sec. III C, where $\mathrm{TE}_{Y\to X}$ identifies Eq. (19) with $\tau \to \infty$).

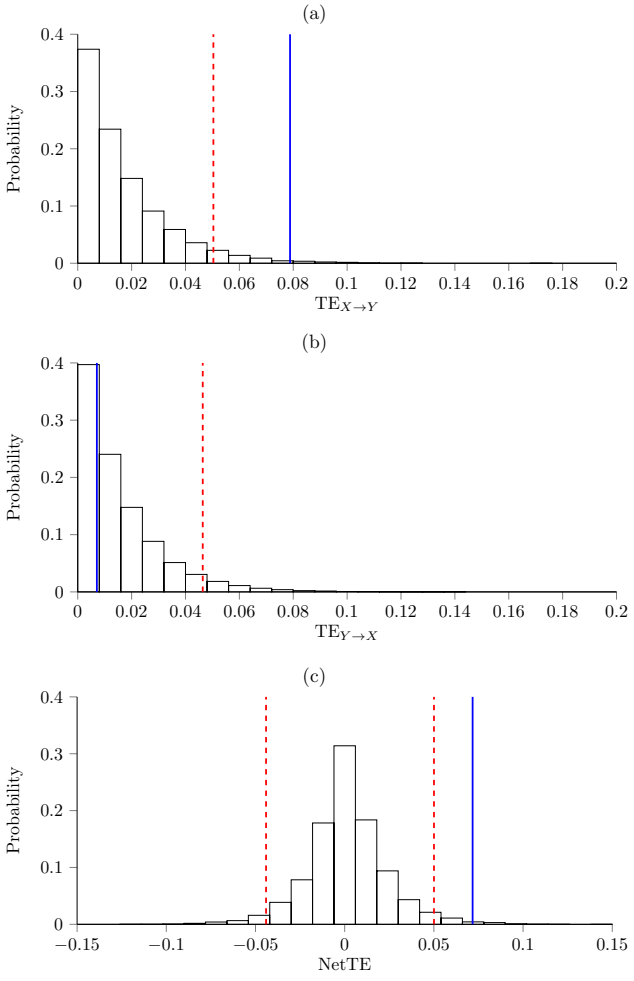

FIG. 3. Distribution of transfer entropy for surrogate data and real transfer entropy: (a) media output to background checks, (b) background checks to media output, and (c) net transfer entropy from media output to background checks. Histograms represent the distribution of transfer entropies over 20 000 surrogate data generated with the IAAFT algorithm, and the blue solid line is the value of transfer entropy for the real datasets. In (a) and (b), the red dashed lines represent the 95% quantile of the distribution, while in (c), they delimit the 2.5% and the 97.5% quantiles.

Consistent with Porfiri *et al.*,[21] from Fig. 3(a) we reject the null hypothesis in ($H1$), identifying a causal link from media output to background checks. This confirms our expectation regarding the role of media coverage over firearm acquisition, whereby people fearing more stringent firearm control laws tend to purchase them before their enactment. The results for transfer entropy from background checks to media output, shown in Fig. 3(b), lead to accepting the null of ($H2$), such that there is no significant information transfer from the former to the latter. In fact, one would not expect that the number of background checks influences media output on legislative aspects, which may be driven by other explanatory factors, such as episodes of mass violence, political elections, and social turmoil. The net transfer entropy in Fig. 3(c) confirms the intuition that the interaction between the two time-series is dominated by a causal link from media output to background checks.

IV. APPLICATION TO THREE CONTROLLED EXPERIMENTAL DATASETS

In this section, we examine the reliability of our approach on three experimental datasets dealing with three types of interactions: fluid-structure,[22] predator-prey,[23] and human-computer.[24] All of these time-series have been analyzed by means of either classical or symbolic transfer entropy in previous publications. In these examples, causal links were experimentally controlled, so that they could serve as valid benchmarks for testing information-theoretic methods for the discovery of causal links.

A. Fluid-structure interaction

First of all, we consider the dataset from Zhang *et al.*,[22] where the authors investigated information transfer between two tandem flapping airfoils in a water tunnel. While this configuration is well-known in the fluid dynamics literature and can be studied with standard analytical, numerical, and experimental techniques, it constitutes a solid basis for validating the use of information-theoretic approaches to study the fluid-structure interaction. Building upon this study, one could focus on more complex instances of fluid-mediated coupling, such as schooling of fish,[36] where there is a dire need of a model-free approach to examine influence.

The time-series of pitch angles were acquired at 30 Hz, for a duration of approximately five minutes, totaling around 9 000 time instants in each time-series. The directionality of the information transfer between the airfoils was controlled by actuating either the upstream or the downstream airfoil while leaving the other free to rotate. The actuated airfoil performed an information-rich pitching motion, whereby it shifted between -15° and 15° following a Markovian model. Overall, 10 trials for each experimental condition were considered, varying the roles of the airfoils (upstream actuated and downstream passive, and vice versa) and their separation distance. As one would expect, paired t-tests on symbolic transfer entropy showed that, for small separation distances, the actuated airfoil influenced the passive airfoil. The extent of the influence, measured through net transfer entropy, diminished when the downstream airfoil was active, whereby the physical pathway of convection could not be utilized for conveying information. By increasing the separation distance, the authors observed a progressive decrease in the influence.

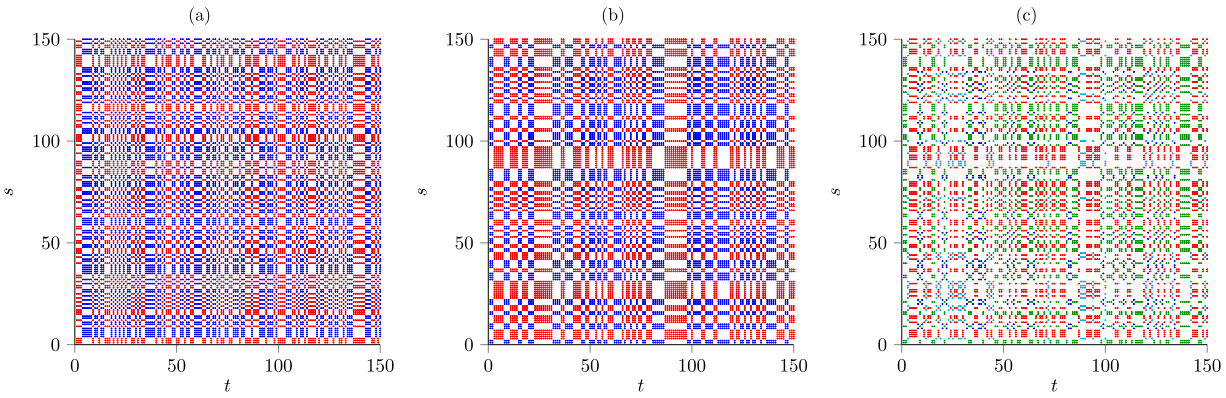

FIG. 4. SRPs of (a) upstream airfoil angle, (b) downstream airfoil angle, and (c) their joint distribution. In (a) and (b), blue and red dots indicate decreasing (π_1) and increasing (π_2) ordinal patterns, respectively. In (c), blue, red, dark green, and cyan indicate decreasing-decreasing (π_1^x, π_1^y), decreasing-increasing (π_1^x, π_2^y), increasing-decreasing (π_2^x, π_1^y), and increasing-increasing (π_2^x, π_2^y) ordinal patterns for the joint symbols, respectively.

TABLE II. SRQA of the datasets on three types of interactions: fluid-structure (a), predator-prey (b), and human-computer (c), illustrating the measures defined in Sec. II B. Here, *X* and *Y* indicate, respectively, in (a) the upstream and downstream airfoil angle, in (b) the robotic predator angular speed and live zebrafish speed, and in (c) the absolute score difference and hand speed, and (*X*, *Y*) represent their joint for each case. The entries of the table represent the mean over the realizations and the standard deviation, in parentheses.

(a)

	$SRR(\pi_1)$	$SRR(\pi_2)$	$d_{\max}$	$\bar{d}$	D	$\bar{v}$
X	0.265 (0.012)	0.235 (0.011)	18.30 (2.111)	3.093 (0.019)	0.385 (0.004)	2.482 (0.040)
Y	0.252 (0.018)	0.248 (0.018)	18.90 (1.101)	3.120 (0.015)	0.370 (0.002)	2.804 (0.060)

	$SRR(\pi_1^x, \pi_1^y)$	$SRR(\pi_1^x, \pi_2^y)$	$SRR(\pi_2^x, \pi_1^y)$	$SRR(\pi_2^x, \pi_2^y)$	$d_{\max}$	$\bar{d}$	D	$\bar{v}$
(X, Y)	0.041 (0.007)	0.097 (0.009)	0.097 (0.009)	0.035 (0.004)	11.60 (1.075)	2.545 (0.018)	0.130 (0.003)	2.251 (0.048)

(b)

	$SRR(\pi_1)$	$SRR(\pi_2)$	$d_{\max}$	$\bar{d}$	D	$\bar{v}$
X	0.190 (0.075)	0.337 (0.126)	62.63 (53.03)	3.400 (0.536)	0.410 (0.059)	3.615 (1.594)
Y	0.251 (0.018)	0.249 (0.019)	19.00 (2.204)	3.081 (0.034)	0.384 (0.005)	2.523 (0.079)

	$SRR(\pi_1^x, \pi_1^y)$	$SRR(\pi_1^x, \pi_2^y)$	$SRR(\pi_2^x, \pi_1^y)$	$SRR(\pi_2^x, \pi_2^y)$	$d_{\max}$	$\bar{d}$	D	$\bar{v}$
(X, Y)	0.048 (0.020)	0.047 (0.018)	0.086 (0.040)	0.083 (0.026)	11.37 (3.249)	2.420 (0.077)	0.127 (0.023)	2.240 (0.132)

(c)

	$SRR(\pi_1)$	$SRR(\pi_2)$	$d_{\max}$	$\bar{d}$	D	$\bar{v}$
X	0.010 (0.003)	0.811 (0.029)	51.69 (9.147)	6.310 (1.016)	0.796 (0.032)	10.35 (2.101)
Y	0.269 (0.012)	0.232 (0.011)	21.37 (3.113)	3.063 (0.028)	0.373 (0.002)	2.682 (0.121)

	$SRR(\pi_1^x, \pi_1^y)$	$SRR(\pi_1^x, \pi_2^y)$	$SRR(\pi_2^x, \pi_1^y)$	$SRR(\pi_2^x, \pi_2^y)$	$d_{\max}$	$\bar{d}$	D	$\bar{v}$
(X, Y)	0.003 (0.001)	0.002 (0.001)	0.212 (0.013)	0.193 (0.013)	16.73 (3.880)	2.718 (0.047)	0.265 (0.015)	2.561 (0.110)

Herein, we focus on the condition with actively controlled upstream airfoil and passive downstream airfoil, a separation distance of 2.5 cm, and a time lapse between actuation episodes of 0.5 s. We downsample the time-series at a resolution of 14 (that is, we skip 13 out of 14 samples), which maximizes net transfer entropy over symbolic recurrence for eight of the ten realizations. As a result, each of the time-series is about 650 samples long, and the duration of a time step is 0.47 s. Due to the availability of multiple time-series, permutation tests are performed to generate surrogate data.

In Fig. 4, SRPs of the time-series of one of the trials are shown for an embedding dimension $m = 2$. For clarity, we select 150 consecutive points, corresponding to 70 s of the experiment. Figure 4(a) illustrates the SRP of the pitch angle of the actuated upstream airfoil. The plot shows few structures with a behavior that approaches that of noise. This is due to the fact that the resolution is only slightly smaller than the time in between actuation events, such that the SRP virtually captures the random variable used to generate the information-rich pitching motion of the active airfoil. The SRP of the pitch angle of the downstream airfoil in Fig. 4(b) shows a more ordered pattern, with wide red bands, up to 5.13 s, of increase in the pitch angle (symbol π_2^y). These bands correspond to a drift in the downstream airfoil that could be associated with friction at the airfoil joint and imperfections of the setup.

The joint SRP in Fig. 4(c) indicates an even more erratic behavior. This plot is dominated by red (π_2^x, π_1^y) and dark green (π_1^x, π_2^y) points, suggesting that the two airfoils tend to move in opposite directions. Physically, this may be explained by the effect of the wake of the upstream airfoil, which causes a lower pressure region on the side of the downstream airfoil aligned with the trailing edge of the upstream airfoil, thereby triggering its motion in the opposite direction.

SRQA of this dataset is presented in Table II(a), where we list means and standard deviations of measures computed over the entire set of downsampled time-series. For both the upstream and downstream airfoils, we observe a substantial similarity between the recurrence rates of each symbol, consistent with the fact that pitching angles should not have a bias in trials. Analogously, the measures based on diagonals' length do not show a considerable difference for the two airfoils, with mean values only slightly higher for the downstream airfoil. On the other hand, the trapping time is higher for the passive downstream airfoil than for the actuated upstream airfoil, indicating an inertia in the change of direction, consistently from what has been observed in the SRP in Fig. 4(b). As expected from Fig. 4(c), the *SRR* for the joint SRP is dominated by the red and dark green symbols, (π_1^x, π_2^y) and (π_2^x, π_1^y), respectively, which have a more than doubled recurrence rate compared to the remaining symbols. The more irregular behavior of the joint SRP is supported by its considerably lower measures based on diagonal and vertical lines, compared to the SRPs of the single airfoils.

Figure 5 shows the results from the symbolic recurrence transfer entropy analysis. As one would expect, based on the results in Fig. 5(a), we reject the null hypothesis in ($H1$), confirming the causal link from the upstream to the downstream airfoil. On the other hand, results in Fig. 5(b) do not allow to reject the null in ($H2$). In agreement with our expectations, we do not register an influence of the downstream airfoil on the upstream one, whose pitching was determined by the commanded actuation. Finally, permutation tests on

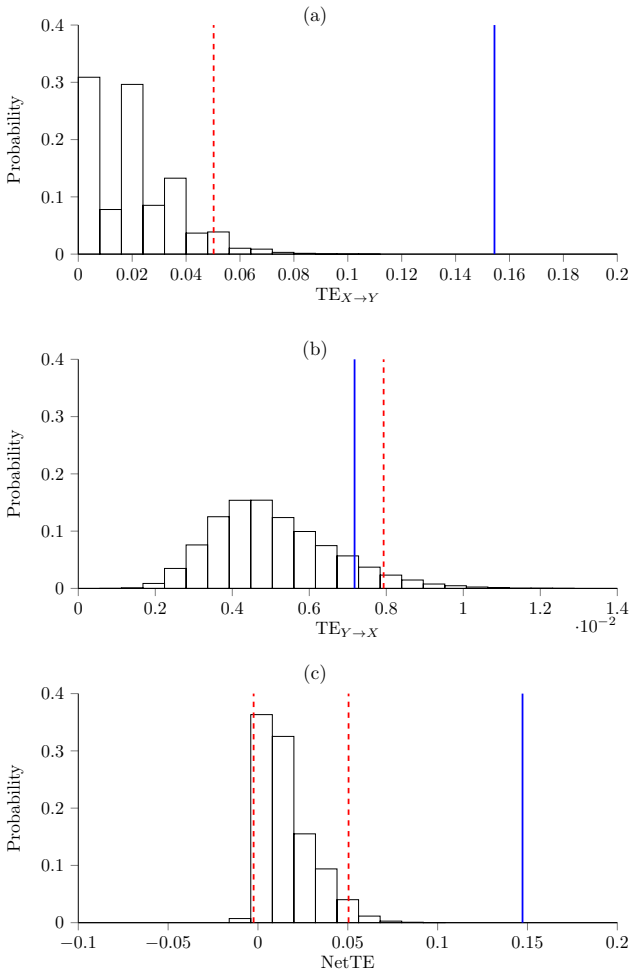

FIG. 5. Distribution of transfer entropy for surrogate data and real transfer entropy: (a) upstream to downstream airfoil angle, (b) downstream to upstream airfoil angle, and (c) net transfer entropy from upstream to downstream airfoil angle. Histograms represent the distribution of transfer entropies over 20 000 surrogate data generated from mean transfer entropies with data shuffling, and the blue solid line is the value of transfer entropy for the real datasets. In (a) and (b), the red dashed lines represent the 95% quantile of the distribution, while in (c), they delimit the 2.5% and the 97.5% quantiles.

net transfer entropy in Fig. 5(c) suggest a significant net information transfer from the upstream to the downstream airfoil, such that we conclude that there exists a fluid-mediated directional influence between the pitching motions of the airfoils.

B. Predator-prey interaction

Here, we reevaluate symbolic recurrence analysis to study animal behavior, extending the analysis proposed in Porfiri and Ruiz Marín[11] on sociality to encompass fear response. Specifically, we examine the dataset from Neri *et al.*[23] that investigated the response

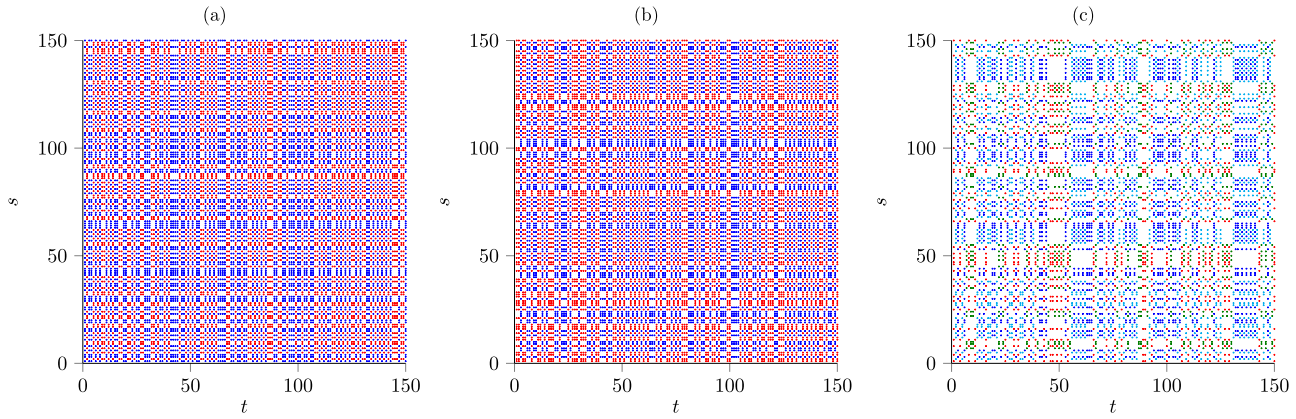

FIG. 6. SRPs of (a) robotic predator angular speed, (b) live zebrafish speed, and (c) their joint distribution. In (a) and (b), blue and red dots indicate decreasing (π_1) and increasing (π_2) ordinal patterns, respectively. In (c), blue, red, dark green, and cyan indicate decreasing-decreasing (π_1^x, π_1^y), decreasing-increasing (π_1^x, π_2^y), increasing-decreasing (π_2^x, π_1^y), and increasing-increasing (π_2^x, π_2^y) ordinal patterns for the joint symbols, respectively.

of zebrafish to live and robotic fear-evoking stimuli. In their experiments, the authors utilized a circular arena, containing a circular Plexiglas tank, partitioning the experimental setup in an inner tank and an outer annulus. Predatory stimuli (a live red tiger oscar fish or a 3D-printed replica of such a fish maneuvered through a cam-based robotic platform) were placed in the inner tank. Zebrafish were allowed to swim in the outer annulus, while visually interacting with the stimuli through the transparent Plexiglas partition. The authors employed an information-theoretic approach based on classical transfer entropy to study the time-series of the predator and zebrafish position.

Experiments in Neri *et al.*[23] consisted of four conditions. Two served as control conditions, whereby either the inner tank was empty or the cam was actuated within the inner tank without a replica. The other two conditions comprised the presentation of a live predator or the 3D-printed replica mounted on the cam-based robotic platform. In the latter condition, the robotic platform performed a start-and-stop motion on a predetermined trajectory, consisting of a sequence of movements at random angular speeds and distances, separated by random time lapses. The parameters of this motion were adjusted to mimic a live predator, utilizing results from pilot tests on live red tiger oscar fish swimming in the inner tank. In each condition, 10 zebrafish were individually tested, and their interaction with the predatory stimuli was recorded with a camera at an acquisition rate of 30 Hz for 21 min. Of these 21 min, the last 10 min were considered for the analysis to account for habituation. The interaction between the predatory stimuli and zebrafish was quantified by binning the circular tank into circular sectors. To tease out causal links between the predator and the prey, the authors computed transfer entropy between the positions of the predator and zebrafish, utilizing different resolutions. From the analysis of net transfer entropy, a causal link from the replica to zebrafish was identified through *t*-tests, whereas no causal links were revealed in the other conditions, for any considered number of bins and resolution.

In this paper, we focus on the interaction between the robotic predator and the zebrafish, to establish whether the robotic predator was effective in influencing zebrafish behavior. Different from Neri *et al.*,[23] we consider the angular speed of the replica, which is forced to move along a circle, and the speed of the zebrafish, which can swim in two dimensions, as the time-series for the analysis. First, we identify outliers from the fish speed through the interquartile range × 1.5 rule.[38] Two of the trials, in which the live zebrafish froze throughout the experiment, are excluded from the analysis. Then, the remaining time-series are downsampled at 1 Hz to mitigate the effect of noise and time-delays between the predator and the zebrafish. To infer causal influence between the predator and the zebrafish, we generate surrogate data from permutation tests, exploiting the availability of multiple time-series.

SRPs of one of the downsampled trials from the current dataset are shown in Fig. 6 for an embedding dimension of $m = 2$. In these plots, we present for clarity only 150 points, corresponding to two and a half minutes of the trial. From Fig. 6(a), we do not find any significant structure in the SRP of the angular speed of the replica. The nature of the start-and-stop motion performed by the robotic platform, which was tuned to replicate the motion of a live predator, may explain this observation. Similarly, the SRP of the live zebrafish speed in Fig. 6(b) does not contain any significant structure, as one might expect from the typical burst-and-coast swimming style of zebrafish in placid water.[39]

The joint SRP in Fig. 6(c) presents some interesting colored patterns, pointing at two different types of response of zebrafish. On the one hand, we observe clusters of blue and cyan dots, corresponding to the symbols (π_1^x, π_1^y) and (π_2^x, π_2^y), respectively, which represent instances in which the replica and the zebrafish simultaneously accelerate or decelerate. On the other hand, smaller blocks of red and green dots, associated with (π_1^x, π_2^y) and (π_2^x, π_1^y), respectively, indicate a response of the zebrafish opposite to the presented stimulus, whereby it accelerates as the predator decelerates and vice versa.

The SRQA of this dataset is presented in Table II(b), where we list the mean and standard deviation of each of the measures discussed in Sec. II B across the eight trials analyzed herein. Symbolic

recurrence rates of the angular speed of the replica present a drastic difference between the two symbols. This disparity can be explained by periods of inactivity (accounted for by the increasing symbol π_2^x) of the replica, which mimics the relatively low activity of live predators in pilot tests.[23] Consistent with this explanation, the mean maximum diagonal length across the trials spans more than 60 s, with a considerable variability among them. These long diagonals correspond to instances of inactivity of the robotic predator in the inner tank. While the value of determinism is high, the average diagonal and vertical lines' length have relatively small values, indicating a disordered behavior as that presented in the SRP in Fig. 6(a).

The symbolic recurrence rates for the speed of zebrafish indicate a more balanced scenario between the two symbols, whereby they have almost the same probability of recurrence. The maximum diagonal length for this variable spans on average 19 s, suggesting a less predictable behavior of live zebrafish compared to the robotic predator. The other measures based on diagonal and vertical lines confirm this observation, whereby their value is consistently smaller than that computed for the angular speed of the replica.

Consistent with these results, recurrence rates for the symbols (π^x, π^y) in the joint process are primarily determined by the corresponding symbol π^x of the replica. On the other hand, these recurrence rates do not show an appreciable dependence on the symbol π^y of zebrafish. While we can visually identify clusters of points of two different colors in the joint SRP in Fig. 6(c) as pointed out above, they do not beget appreciable measures in the SRQA of the joint. The measures on diagonal and vertical lines for the joint, in fact, have significantly lower values compared to those for the single processes.

In order to elucidate whether the robotic predator influences the behavior of the live zebrafish, we perform statistical tests on transfer entropy on symbolic recurrences. The results of transfer entropy are shown in Fig. 7. From Fig. 7(a), we reject the null hypothesis in ($H1$), thereby identifying an influence of the replica on zebrafish. As one would expect from the predetermined motion of the cam-based platform, results in Fig. 7(b) do not indicate a causal link from zebrafish to the replica, accepting the null in ($H2$). Consistent with these two claims, permutation tests on net tranfer entropy in Fig. 7(c) indicate that the angular speed of the replica has a net influence on zebrafish speed, supporting the use of robots in animal behavior to reduce the number of subjects utilized in the experiments and provide consistent stimuli across trials.[40,41]

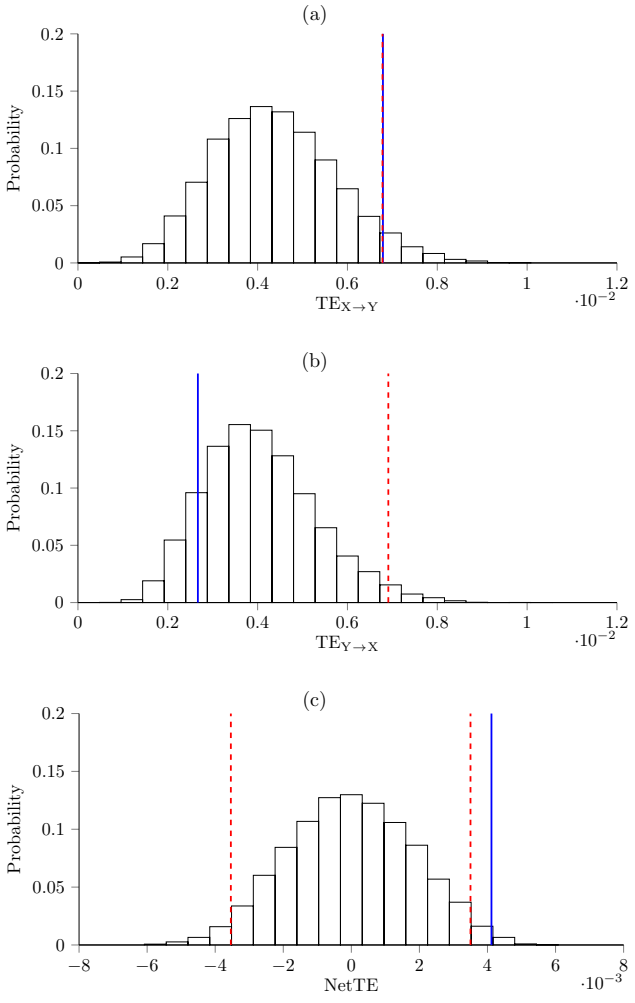

FIG. 7. Distribution of transfer entropy for surrogate data and real transfer entropy: (a) robotic predator angular speed to live zebrafish speed, (b) live zebrafish speed to robotic predator angular speed, and (c) net transfer entropy from robotic predator angular speed to live zebrafish speed. Histograms represent the distribution of transfer entropies over 20 000 surrogate data generated from mean transfer entropies with data shuffling, and the blue solid line is the value of transfer entropy for the real datasets. In (a) and (b), the red dashed lines represent the 95% quantile of the distribution, while in (c), they delimit the 2.5% and the 97.5% quantiles.

C. Human-computer interaction

Finally, we consider experiments in Barak Ventura *et al.*[24] as an example of time-series related to human behavior. In the paper, the authors sought to identify whether there is an influence of winning or losing on the engagement of players in a virtual reality (VR) game. In the game, each player competed for points against an under-performing, equally-performing, or over-performing virtual opponent. The game consisted in popping floating balloons that appeared randomly in the VR environment using a hand controller. While players were deceived to believe that the opponent was another human player, the virtual opponent was programed to simulate the different conditions systematically to help tease out the role of competition on engagement. During each trial, the absolute score difference between the player and the virtual opponent, along with the hand speed, was acquired at 100 Hz. The absolute score difference should be considered as a measure of players' knowledge regarding their relative performance, providing a measure of the level of competition in the game. On the other hand, the hand speed should quantify engagement of the player, as previously considered in the literature of behavioral studies.[42]

In their experiments, Barak Ventura *et al.*[24] analyzed 72 trials with volunteers. From classical symbolic transfer entropy analyses,

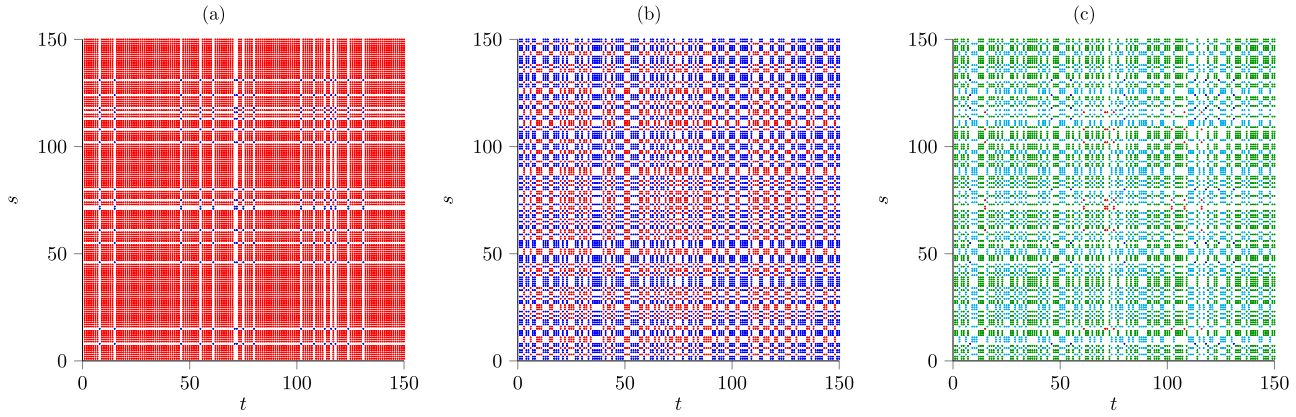

FIG. 8. SRPs of (a) absolute score difference, (b) hand speed, and (c) their joint distribution. In (a) and (b), blue and red dots indicate decreasing (π_1) and increasing (π_2) ordinal patterns, respectively. In (c), blue, red, dark green, and cyan indicate decreasing-decreasing (π_1^x, π_1^y), decreasing-increasing (π_1^x, π_2^y), increasing-decreasing (π_2^x, π_1^y), and increasing-increasing (π_2^x, π_2^y) ordinal patterns for the joint symbols, respectively.

the authors concluded that players are more engaged when competing against an over-performing opponent, such that they would tend to lose. This effect may be explained through the lens of flow theory,[43] which suggests that engagement reduces in extreme cases of tasks that are excessively easy or challenging, unlike the proposed VR game where they lose by only a few points.

Here, we consider only such an over-performing condition, using as time-series the absolute score difference (ASD) and the hand speed (HS). HS time-series have some isolated missing data points, due to acquisition problems. To overcome this issue, we utilize as a value for the missing points the linear interpolation between the two neighboring data points (this strategy is automatically adopted in the provided software). Following Barak Ventura *et al.*,[24] we downsample at a resolution of 12 to reduce the effect of noise, but we consider symbolization with ordinal patterns and embedding dimension $m = 2$ for both variables, instead of three symbols for the ASD as in the original paper.

In Fig. 8, we show the SRPs of the time-series from one of the trials, again limited to 150 time instants for the sake of clarity. The SRP of the absolute score difference in Fig. 8(a) shows a net prevalence of the red symbol (π_2^x), which can be explained by considering the nature of the variable. In fact, ASD changes only when the player or the virtual opponent perform an action. Specifically, the blue symbols (π_1^x) indicate a decrease in the score difference, corresponding to either the winner losing a point or the loser gaining one, while the red symbols (π_2^x) represent instances in which the gap remains the same or increases. The red bands in the SRP indicate instances of reduced activity of the human player, suggesting exploration of the virtual space that is interrupted by isolated actions when the human player pops the balloons. Hand speed, instead, has a less structured SRP [Fig. 8(b)]. As one would expect, the motion of the hand is more erratic, with rapid increases and decreases of the speed, typical of computer games. The joint SRP in Fig. 8(c) inherits this disorganized pattern, with a net prevalence of the dark green (π_2^x, π_1^y) and cyan (π_2^x, π_2^y) symbols, associated with the increasing symbol π_2^x for the ASD.

These observations are reflected in the SRQA in Table II(c), which shows means and standard deviations computed from all the trials. The SRP of the ASD shows predominance of the red symbol (π_2^x), which covers more than 80% of the SRP. On the other hand, the SRP of HS presents a substantial equilibrium in the recurrence of each symbol, with recurrence points covering on average half of the SRP. In agreement with our observations on the structures of their SRPs, the maximum diagonal length, average diagonal length, and determinism are higher for the absolute score difference than for the hand speed. The average vertical length also follows this trend, confirming the more structured nature of ASD data. The measures on the joint SRP illustrate a prevalence of the joint symbols that contain π_2^x, irrespective of HS, with lower values for both diagonal and vertical lines, which indicate increasing disorder in the joint SRP.

Similar to Barak Ventura *et al.*,[24] we test only whether there is a causal link from the absolute score difference to the hand speed,

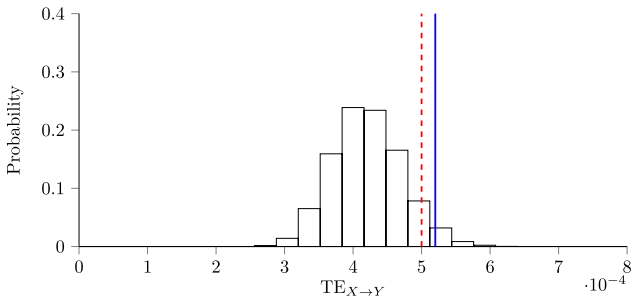

FIG. 9. Distribution of transfer entropy for surrogate data and real transfer entropy from absolute score difference to hand speed. Histograms represent the distribution of transfer entropies over 20 000 surrogate data generated from mean transfer entropies with data shuffling, and the blue solid line is the value of transfer entropy for the real datasets. The red dashed line represents the 95% quantile of the distribution.

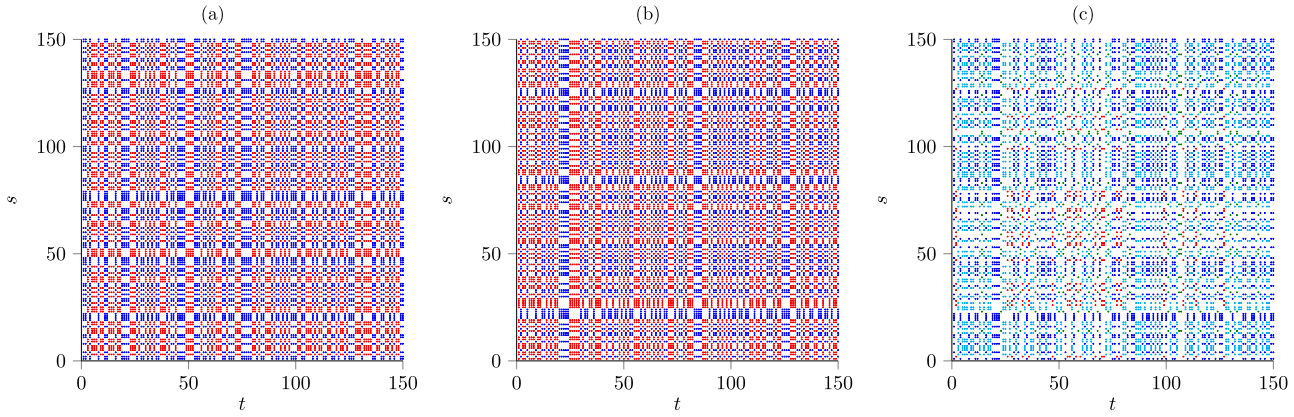

FIG. 10. SRPs of (a) male zebrafish speed, (b) female zebrafish speed, and (c) their joint distribution. In (a) and (b), blue and red dots indicate decreasing (π_1) and increasing (π_2) ordinal patterns, respectively. In (c), blue, red, dark green, and cyan indicate decreasing-decreasing (π_1^x, π_1^y), decreasing-increasing (π_1^x, π_2^y), increasing-decreasing (π_2^x, π_1^y), and increasing-increasing (π_2^x, π_2^y) ordinal patterns for the joint symbols, respectively.

that is, we only test hypothesis ($H1$) to unravel the effect of competition on engagement. Due to the availability of multiple time-series, we conduct a permutation test, whose results are displayed in Fig. 9. Consistent with our expectation and the results from the original paper, we find an influence of the absolute score difference on the hand speed, demonstrating that the engagement of the player is influenced by his performance against a winning virtual opponent.

V. EXPLORING THE ROLE OF SEX ON SOCIAL INTERACTION IN FISH

Here, we demonstrate the application of the proposed framework to help answer scientific questions from time-series, without *a priori* knowledge about potential causal links underlying the dataset. Specifically, we seek to tease out the role of sex in pairs of zebrafish swimming together. The dataset was originally presented in Neri *et al.*,[25] as part of pilot experiments to elucidate whether sex had an effect on leadership, measured as the tendency of fish to initiate new directions in the motion that are readily followed by other members of the group.[44] Such an analysis was partially motivated by relevant literature suggesting that sex of zebrafish might influence their shoaling tendency,[45] whereby males had a higher preference to shoal with females, whereas the preference of females for other fish did not vary with their sex.

In Neri *et al.*,[25] a male and a female zebrafish were allowed to swim in the same tank for about 15 min, and the interaction was recorded with a camera at 40 Hz. The first 10 min of the experiment were treated as habituation, and they were discarded from the analysis. Trajectories of individual fish in the remaining 5 min were obtained using a custom-made tracking software and then processed to extract the velocity of each fish. Overall, 14 pairs of zebrafish were considered, but three pairs were discarded, two for technical problems and one after testing for outliers. Such a dataset was not originally analyzed within an information-theoretic perspective. Rather, the authors implemented classical correlation tools, where leadership was inferred by evaluating the time-lags between fish that maximized the value of the cross correlation between their swimming directions. In this sense, the analysis identifies which fish is mirroring the other and which is the delay between them. Neither two-tailed t-tests on the distributions of the time-lags for males and females nor χ^2-tests on the frequency by which males and females act as leaders support a significant effect of sex on leadership.

TABLE III. SRQA of the dataset on the role of sex on social interactions in zebrafish, illustrating the measures defined in Sec. II B. Here, X indicates the male zebrafish speed, Y refers to the female zebrafish speed, and (X, Y) is their joint process. The entries of the table represent the mean over the realizations and the standard deviation, in parentheses.

	$SRR(\pi_1)$		$SRR(\pi_2)$		$d_{\max}$	$\bar{d}$	D	$\bar{v}$
X	0.250		0.249		15.64	3.078	0.378	2.503
	(0.010)		(0.010)		(2.976)	(0.075)	(0.008)	(0.126)
Y	0.268		0.232		15.18	3.063	0.383	2.531
	(0.028)		(0.026)		(1.601)	(0.036)	(0.004)	(0.119)
	$SRR(\pi_1^x, \pi_1^y)$	$SRR(\pi_1^x, \pi_2^y)$	$SRR(\pi_2^x, \pi_1^y)$	$SRR(\pi_2^x, \pi_2^y)$	$d_{\max}$	$\bar{d}$	D	$\bar{v}$
(X, Y)	0.104	0.032	0.039	0.093	8.727	2.414	0.132	2.246
	(0.014)	(0.009)	(0.014)	(0.021)	(1.191)	(0.029)	(0.011)	(0.095)

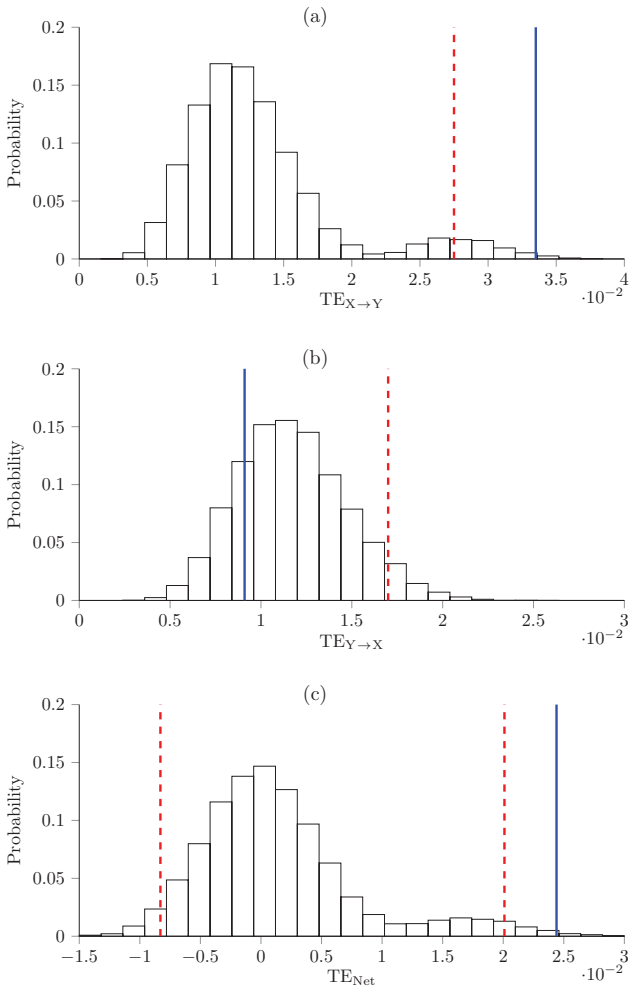

FIG. 11. Distribution of transfer entropy for surrogate data and real transfer entropy: (a) male zebrafish speed to female zebrafish speed, (b) female zebrafish speed to male zebrafish speed, and (c) net transfer entropy. Histograms represent the distribution of transfer entropies over 20 000 surrogate data generated from mean transfer entropies with data shuffling, and the blue solid line is the value of transfer entropy for the real datasets. In (a) and (b), the red dashed lines represent the 95% quantile of the distribution, while in (c), they delimit the 2.5% and the 97.5% quantiles.

Here, we perform symbolic recurrence analysis on this dataset, considering the speed of the two fish as the variables in the analysis. In order to reduce the effect of noise, we downsample the time-series to 1 Hz. This value is selected to maximize the absolute value of the net transfer entropy between the time-series, mirroring the analysis in Sec. IV A. The availability of multiple realizations allows us to utilize permutation tests to generate surrogate data.

An example of the individual SRPs of the speed time-series of male and female zebrafish along their joint SRP is shown in Fig. 10. Again, we utilize an embedding dimension of $m = 2$, and we depict only 150 points from the downsampled time-series. As one might expect from zebrafish swimming style in placid water, SRPs of the speed of both the male and female subjects in Figs. 10(a) and 10(b) do not contain any relevant structure and simply resemble noise. Interestingly, while the joint SRP in Fig. 10(c) inherits the lack of structure from the SRPs of the single variables, it is mostly populated by blue and cyan dots, corresponding to the symbols (π_1^x, π_1^y) and (π_2^x, π_2^y).

SRQA of this dataset is summarized in Table III, where we report means and standard deviations of the measures proposed in Sec. II B for the entire dataset of 11 fish pairs. First of all, we observe that *SRR*s of males present identical behavior in terms of acceleration and deceleration over time, while SRPs of females display a slight prevalence of acceleration instances. As expected from the SRPs, measures based on diagonal and vertical lines do not indicate a difference between males and females, whereby their behavior does not reflect any predetermined structure.

Consistent with observations on the joint SRP in Fig. 10(c), we find that the recurrence rate of the symbols (π_1^x, π_1^y) and (π_2^x, π_2^y) is more than double the recurrence rate of the other two symbols. This indicates that male and female zebrafish tend to accelerate and decelerate simultaneously while swimming together, suggesting some form of social interaction between them. However, measures on diagonal and vertical lines have relatively small values, in agreement with the erratic behavior observed in the joint SRP.

Transfer entropy on symbolic recurrences is computed to elucidate sex-related differences in social behavior, see Fig. 11. Based on results in Fig. 11(a), we reject the null hypothesis in ($H1$), and we propose that the swimming patterns of males influence the swimming patterns of females. Interestingly, Fig. 11(b) suggests that this influence is not reciprocated, whereby we accept the null in ($H2$) that there is not an influence of females on males. The results of permutation tests on net transfer entropy, shown in Fig. 11(c), yield the rejection of the null hypothesis in ($H3$) and confirm the presence of a net influence of males on females.[46]

This causal link from male to female zebrafish opens up several interpretations from a biological perspective. A possible explanation relates to sexual harassment, a behavior that has been widely documented in zebrafish.[47–49] In this case, our results would suggest a situation in which males harass females through repeated approaches, so that females seek to escape from the males' attempts to mate. This interpretation would also explain the propensity of both fish to accelerate or decelerate simultaneously. On the other hand, we would expect short-term interactions rather than chasing behavior of the males over an extended period of time, which would otherwise be influenced by females. Another possibility is based on higher risk-taking behavior of male zebrafish compared to females,[50] which might have triggered some form of leadership[51] that mere correlation analysis could not detect. In this vein, males would be more prone to explore the tank, with females following them. While the correlation analysis assumes that a follower must copy the behavior of a leader, an information-theoretic perspective can accommodate more general nonlinear interactions that might be at the basis of the leader–follower relationship.[52]

VI. CONCLUSIONS

Recurrence plots and recurrence quantification analysis are widely used to investigate the behavior of dynamical systems from

their raw time-series. However, in their traditional incarnation, they are affected by the selection of a proximity parameter ε in the phase space, which introduces arbitrariness in their construction. Recurrence on symbolic dynamics allows for overcoming this problem, thereby establishing a parameter-free approach that goes beyond simple black-and-white representations of dynamical systems. Symbolic recurrence plots offer a colored representation of recurrence, visualizing the portions of the phase space that are repeatedly visited by the system. From the dynamics of the symbols, one can define symbolic recurrence quantification analysis and transfer entropy on symbolic recurrences, creating a powerful toolbox for time-series analysis.

This paper contributes to the state-of-the-art of recurrence methods by demonstrating the potential of symbolic recurrence plots, symbolic recurrence quantification analysis, and transfer entropy on symbolic recurrences toward the unprecedented analysis of real time-series. In this vein, we have tested the validity of these tools on several datasets across different areas in mechanics and behavioral sciences, illustrating their implementation and highlighting their technical merit.

While transfer entropy on symbolic recurrences is able to identify causal links between dynamical systems, symbolic recurrence plots and symbolic recurrence quantification analysis provide a complementary perspective to unfold key features of dynamical systems. For example, in the study of the firearm prevalence dataset,[21] symbolic recurrence plots have helped in pinpointing key events that shaped the current firearm ecosystem. From the election of President Obama in 2008 to the Sandy Hook mass shooting in 2012, a number of key events have increased people's fear of more stringent firearm regulations, prompting them to seek to purchase new firearms.

A particularly enticing advantage of this framework entails its colored patterns. Colors in the symbolic recurrence plots indicate the portion of the phase space that the dynamical system is repeatedly visiting, providing important insight into the system that depends on the choice of the phase space partition. For example, when using an embedding dimension of two and ordinal patterns, these colors identify instances of increase and decrease of the variable of interest between two consecutive points in the time-series.

This insight, absent in classical black-and-white recurrence plots, could help identify the form of the relationship between dynamical systems. For instance, in the joint process in the fluid-structure interaction dataset,[22] the higher recurrence rate of increasing-decreasing (π_2^x, π_1^y) and decreasing-increasing (π_1^x, π_2^y) symbols enables the identification of a physical pathway of interaction, through the wake of the upstream airfoil.

In some cases, this insight may be fundamental in teasing out the correct interpretation of transfer entropy results. When considering the role of sex on the social behavior of zebrafish,[25] we have observed that symbols in the joint process associated with increasing-increasing (π_2^x, π_2^y) and decreasing-decreasing (π_1^x, π_1^y) speeds recur twice as often as the other two symbols. Thus, fish tend to increase and decrease their speed simultaneously, such that the causal link identified by transfer entropy might relate to sexual harassment or some form of the leader-follower relationship.

While this paper serves as a primer for the use of a symbolic framework in the study of recurrence, further work is required to widen its applicability to more complex scenarios. For example, the present implementation of transfer entropy on symbolic recurrences does not account for time-delays and time-histories.[53] Also, the present approach is limited to dyadic interactions, which could be overcome by extending conditional transfer entropy[53] to symbolic recurrences. Another direction of improvement consists in releasing the assumption of stationarity of the stochastic process, encompassing cyclostationary[54] and switching[55] processes. Overall, we hope that this paper and the associated software will contribute to providing researchers with novel analysis tools to examine dynamical systems from time-series.

SUPPLEMENTARY MATERIAL

See the supplementary material for the freely-available supplementary software (provided as a Matlab application) to obtain SRPs, SRQA, and transfer entropy on symbolic recurrences for arbitrary datasets. A detailed manual and the datasets utilized in this manuscript are provided along with the software.

ACKNOWLEDGMENTS

This work was supported by the National Science Foundation (NSF) under Grant Nos. CMMI 1433670, CMMI 1505832, and CMMI 1901697 and by a Grant from the New York University Research Challenge Fund Program. This study is part of the collaborative activities carried out under the program Groups of Excellence of the region of Murcia, the Fundación Séneca, Science and Technology Agency of the region of Murcia (Spain) (Project No. 19884/GERM/15).

APPENDIX: SUPPLEMENTARY SOFTWARE

To facilitate the use of the tools described in this paper, a freely available software with a user-friendly graphical user interface (GUI) is provided. With this software, users can obtain SRPs and SRQA from the time-series of a single variable and compute transfer entropy on symbolic recurrences from datasets of two variables via ordinal patterns for symbolization. The software is available in the supplementary material of this paper as a Matlab application. A detailed manual accompanies the software, encompassing the step-by-step installation procedure and description of each option in the GUI.

The GUI of the software entails different tabs, which can be freely navigated and selected by the user. The following tabs are included in the GUI: "Initialize," "Symbolic Recurrence Plots," "Symbolic Recurrence Quantification Analysis," and "Transfer Entropy on Symbolic Recurrences."

The Initialize tab encompasses the selection of datasets and general options for the symbolic recurrence analysis. Datasets are imported by selecting .csv, .txt, .xls, or .xlsx files from the computer. Datasets' files require a special formatting, whereby columns of the files must indicate the different realizations of the considered variable. Examples illustrating the required formatting can be shown in a pop-up window. In the GUI, it is possible to import datasets for two variables. Should the user be interested in obtaining SRPs and SRQA for a single variable, he/she can import the same dataset for each of the variables, and disregard the results for the second variable in the

Symbolic Recurrence Plots and Symbolic Recurrence Quantification Analysis tabs. General options for symbolic analyses comprise the embedding dimension, downsampling of the time-series, and activation of parallel computing. From this table, it is possible to open the user manual for the software. Users can test the software with the firearm prevalence data already incorporated into the software by clicking the "Load Sample Data" button.

Within the Symbolic Recurrence Plots tab, users can visualize and export the SRPs of the imported datasets. In the case of two imported variables, the GUI shows the SRP of the single processes and their joint. When multiple realizations are available, it is possible to select the pair of processes to be plotted. In addition, users can choose the initial time instant for the SRPs and the length of the time-series to be considered in the SRPs.

The Symbolic Recurrence Quantification Analysis tab allows users to obtain the measures listed in Sec. II B for the imported time-series. When considering the overall SRP, which does not bookkeeps the identity of the recurring symbols, it is possible to generate the symbolic recurrence rate, maximum diagonal length, average diagonal length, determinism, and average diagonal length of the single variables and of their joint. Additionally, the symbolic recurrence rates of each of the symbols can be computed with the corresponding option. When the datasets include multiple realizations, the software provides the mean and standard deviation across the different time-series.

Finally, the Transfer Entropy on Symbolic Recurrences tab provides the possibility of inferring causal links between two different variables, by plotting and exporting the histograms of empirical null distributions, 5% significance level, and transfers entropy from the real dataset. Users can choose the number of surrogate data to be generated and the method to create them. Specifically, three methods to generate surrogate data are implemented in the software: pairwise shuffling (valid for multiple realizations only), the IAAFT algorithm, and stationary block bootstrap.[56] For the IAAFT algorithm and stationary block bootstrap, it is possible to select the pair from which surrogate data are computed, to ascertain significance for each single pair of time-series at a level of 0.05. The minimum time step between recurrences can be decided by the user.

REFERENCES

[1] H. Poincaré, "Sur le problème des trois corps et les équations de la dynamique," Acta Math. **13**, A3–A270 (1890).
[2] J.-P. Eckmann, S. O. Kamphorst, and D. Ruelle, "Recurrence plots of dynamical systems," Europhys. Lett. **4**, 973 (1987).
[3] N. Marwan, "A historical review of recurrence plots," Eur. Phys. J. Spec. Top. **164**, 3–12 (2008).
[4] N. Marwan, M. C. Romano, M. Thiel, and J. Kurths, "Recurrence plots for the analysis of complex systems," Phys. Rep. **438**, 237–329 (2007).
[5] N. Marwan, M. H. Trauth, M. Vuille, and J. Kurths, "Comparing modern and Pleistocene ENSO-like influences in NW Argentina using nonlinear time series analysis methods," Clim. Dyn. **21**, 317–326 (2003).
[6] U. Brandt, N. Nowaczyk, A. Ramrath, A. Brauer, J. Mingram, S. Wulf, and J. Negendank, "Palaeomagnetism of Holocene and late Pleistocene sediments from Lago di Mezzano and Lago Grande di Monticchio (Italy): Initial results," Quat. Sci. Rev. **18**, 961–976 (1999).
[7] J. Nichols, S. Trickey, and M. Seaver, "Damage detection using multivariate recurrence quantification analysis," Mech. Syst. Signal Process. **20**, 421–437 (2006).
[8] J. Belaire-Franch, D. Contreras, and L. Tordera-Lledó, "Assessing nonlinear structures in real exchange rates using recurrence plot strategies," Physica D **171**, 249–264 (2002).
[9] N. Marwan and A. Meinke, "Extended recurrence plot analysis and its application to ERP data," Int. J. Bifurcat. Chaos **14**, 761–771 (2004).
[10] M. V. Caballero-Pintado, M. Matilla-García, and M. Ruiz Marín, "Symbolic recurrence plots to analyze dynamical systems," Chaos **28**, 063112 (2018).
[11] M. Porfiri and M. Ruiz Marín, "Transfer entropy on symbolic recurrences," Chaos **29**, 063123 (2019).
[12] N. Marwan and J. Kurths, "Nonlinear analysis of bivariate data with cross recurrence plots," Phys. Lett. A **302**, 299–307 (2002).
[13] Y. Zou, M. C. Romano, M. Thiel, N. Marwan, and J. Kurths, "Inferring indirect coupling by means of recurrences," Int. J. Bifurcat. Chaos **21**, 1099–1111 (2011).
[14] M. C. Romano, M. Thiel, J. Kurths, and C. Grebogi, "Estimation of the direction of the coupling by conditional probabilities of recurrence," Phys. Rev. E **76**, 036211 (2007).
[15] C. L. Webber and J. P. Zbilut, "Dynamical assessment of physiological systems and states using recurrence plot strategies," J. Appl. Physiol. Respir. Environ. Exerc. Physiol. **76**, 965–973 (1994).
[16] T. March, S. Chapman, and R. Dendy, "Recurrence plot statistics and the effect of embedding," Physica D **200**, 171–184 (2005).
[17] A. M. Ramos, A. Builes-Jaramillo, G. Poveda, B. Goswami, E. E. Macau, J. Kurths, and N. Marwan, "Recurrence measure of conditional dependence and applications," Phys. Rev. E **95**, 052206 (2017).
[18] J. M. Amigó, R. Monetti, T. Aschenbrenner, and W. Bunk, "Transcripts: An algebraic approach to coupled time series," Chaos **22**, 013105 (2012).
[19] J. M. Amigó, T. Aschenbrenner, W. Bunk, and R. Monetti, "Dimensional reduction of conditional algebraic multi-information via transcripts," Inf. Sci. **278**, 298–310 (2014).
[20] T. Schreiber, "Measuring information transfer," Phys. Rev. Lett. **85**, 461 (2000).
[21] M. Porfiri, R. R. Sattanapalle, S. Nakayama, J. Macinko, and R. Sipahi, "Media coverage and firearm acquisition in the aftermath of a mass shooting," Nat. Human Behav. **3**, 913–921 (2019).
[22] P. Zhang, M. Rosen, S. D. Peterson, and M. Porfiri, "An information-theoretic approach to study fluid–structure interactions," J. Fluid Mech. **848**, 968–986 (2018).
[23] D. Neri, T. Ruberto, G. Cord-Cruz, and M. Porfiri, "Information theory and robotics meet to study predator-prey interactions," Chaos **27**, 73111 (2017).
[24] R. Barak Ventura, S. Richmond, M. Nadini, S. Nakayama, and M. Porfiri, "Does winning or losing change players' engagement in competitive games? Experiments in virtual reality," IEEE Trans. Games (in press).
[25] D. Neri, T. Ruberto, V. Mwaffo, T. Bartolini, and M. Porfiri, "Social environment modulates anxiogenic effects of caffeine in zebrafish," Behav. Pharmacol. **30**, 45–58 (2019).
[26] F. Takens, "Detecting strange attractors in turbulence," in *Dynamical Systems and Turbulence, Warwick 1980* (Springer, 1981), pp. 366–381.
[27] H. Yang and Y. Chen, "Heterogeneous recurrence monitoring and control of nonlinear stochastic processes," Chaos **24**, 013138 (2014).
[28] A. Groth, "Visualization of coupling in time series by order recurrence plots," Phys. Rev. E **72**, 046220 (2005).
[29] S. Schinkel, N. Marwan, and J. Kurths, "Order patterns recurrence plots in the analysis of ERP data," Cogn. Neurodyn. **1**, 317–325 (2007).
[30] R. Donner, U. Hinrichs, and B. Scholz-Reiter, "Symbolic recurrence plots: A new quantitative framework for performance analysis of manufacturing networks," Eur. Phys. J. Spec. Top. **164**, 85–104 (2008).
[31] M. V. Caballero-Pintado, M. Matilla-García, and M. R. Marín, "Symbolic correlation integral," Econom. Rev. **38**, 533–556 (2019).
[32] T. M. Cover and J. A. Thomas, *Elements of Information Theory* (John Wiley & Sons, 2012).
[33] T. Schreiber and A. Schmitz, "Improved surrogate data for nonlinearity tests," Phys. Rev. Lett. **77**, 635 (1996).
[34] A. P. Holmes, R. C. Blair, J. D. G. Watson, and I. Ford, "Nonparametric analysis of statistic images from functional mapping experiments," J. Cerebral Blood Flow Metab. **16**, 7–22 (1996).
[35] T. E. Nichols and A. P. Holmes, "Nonparametric permutation tests for functional neuroimaging: A primer with examples," Hum. Brain Mapp. **15**, 1–25 (2002).

[36]P. Zhang, E. Krasner, S. D. Peterson, and M. Porfiri, "An information-theoretic study of fish swimming in the wake of a pitching airfoil," Physica D **396**, 35–46 (2019).
[37]E. Depetris-Chauvin, "Fear of Obama: An empirical study of the demand for guns and the U.S. 2008 presidential election," J. Public Econ. **130**, 66–79 (2015).
[38]P. William Navidi, *Statistics for Engineers and Scientists* (McGraw-Hill Education, 2014).
[39]U. Muller, E. Stamhuis, and J. Videler, "Hydrodynamics of unsteady fish swimming and the effects of body size: Comparing the flow fields of fish larvae and adults," J. Exp. Biol. **203**, 193–206 (2000).
[40]M. Porfiri, "Inferring causal relationships in zebrafish-robot interactions through transfer entropy: A small lure to catch a big fish," Anim. Behav. Cogn. **5**, 341–367 (2018).
[41]D. Romano, E. Donati, G. Benelli, and C. Stefanini, "A review on animal–robot interaction: From bio-hybrid organisms to mixed societies," Biol. Cybern. **113**, 201–225 (2019).
[42]R. P. McMahan, D. A. Bowman, D. J. Zielinski, and R. B. Brady, "Evaluating display fidelity and interaction fidelity in a virtual reality game," IEEE Trans. Vis. Comput. Graph **18**, 626–633 (2012).
[43]M. Csikszentmihalyi, S. Abuhamdeh, and J. Nakamura, "Flow," in *Flow and the Foundations of Positive Psychology: The Collected Works of Mihaly Csikszentmihalyi* (Springer, 2014), pp. 227–238.
[44]J. Krause, D. Hoare, S. Krause, C. K. Hemelrijk, and D. I. Rubenstein, "Leadership in fish shoals," Fish Fish. **1**, 82–89 (2000).
[45]N. Ruhl and S. P. McRobert, "The effect of sex and shoal size on shoaling behaviour in *Danio rerio*," J. Fish Biol. **67**, 1318–1326 (2005).
[46]Additional analyses have been carried out in order to ensure that the claims of the study are robust to noise. Specifically, we performed the same procedure on a modified time-series with additive Gaussian noise (with zero mean and standard deviation equal to 10% of the average speed among all the fish), obtaining equivalent results.
[47]A. Etinger, J. Lebron, and B. G. Palestis, "Sex-assortative shoaling in zebrafish (*Danio rerio*)," BIOS **80**, 153–158 (2009).
[48]R. Spence and C. Smith, "Male territoriality mediates density and sex ratio effects on oviposition in the zebrafish, *Danio rerio*," Anim. Behav. **69**, 1317–1323 (2005).
[49]S. Uusi-Heikkilä, D. Bierbach, J. Alós, P. Tscheligi, C. Wolter, and R. Arlinghaus, "Relatively large males lower reproductive success in female zebrafish," Environ. Biol. Fishes **101**, 1625–1638 (2018).
[50]T. Roy and A. Bhat, "Population, sex and body size: Determinants of behavioural variations and behavioural correlations among wild zebrafish *Danio rerio*, R. Soc. Open Sci. **5**, 170978 (2018).
[51]A. J. King, D. D. Johnson, and M. V. Vugt, "The origins and evolution of leadership," Curr. Biol. **19**, R911–R916 (2009).
[52]A. Strandburg-Peshkin, D. Papageorgiou, M. C. Crofoot, and D. R. Farine, "Inferring influence and leadership in moving animal groups," Philos. Trans. R. Soc. B **373**, 20170006 (2018).
[53]T. Bossomaier, L. Barnett, M. Harré, and J. T. Lizier, *An Introduction to Transfer Entropy: Information Flow in Complex Systems* (Springer, Berlin, 2016).
[54]M. Porfiri and M. Ruiz Marín, "Inference of time-varying networks through transfer entropy, the case of a Boolean network model," Chaos **28**, 103123 (2018).
[55]S. Butail and M. Porfiri, "Detecting switching leadership in collective motion," Chaos **29**, 011102 (2019).
[56]D. N. Politis and H. White, "Automatic block-length selection for the dependent bootstrap," Econom. Rev. **23**, 53–70 (2004).