UC Davis
UC Davis Previously Published Works

Title
Modes of information flow in collective cohesion

Permalink
https://escholarship.org/uc/item/52d1k286

Journal
Science Advances, 8(6)

ISSN
2375-2548

Authors
Sattari, Sulimon
Basak, Udoy S
James, Ryan G
et al.

Publication Date
2022-02-11

DOI
10.1126/sciadv.abj1720

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

SYSTEMS BIOLOGY

Modes of information flow in collective cohesion

Sulimon Sattari[1], Udoy S. Basak[1,2], Ryan G. James[3,4], Louis W. Perrin[1,5], James P. Crutchfield[4], Tamiki Komatsuzaki[1,6,7]*

Copyright © 2022 The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original U.S. Government Works. Distributed under a Creative Commons Attribution NonCommercial License 4.0 (CC BY-NC).

Pairwise interactions are fundamental drivers of collective behavior—responsible for group cohesion. The abiding question is how each individual influences the collective. However, time-delayed mutual information and transfer entropy, commonly used to quantify mutual influence in aggregated individuals, can result in misleading interpretations. Here, we show that these information measures have substantial pitfalls in measuring information flow between agents from their trajectories. We decompose the information measures into three distinct modes of information flow to expose the role of individual and group memory in collective behavior. It is found that decomposed information modes between a single pair of agents reveal the nature of mutual influence involving many-body nonadditive interactions without conditioning on additional agents. The pairwise decomposed modes of information flow facilitate an improved diagnosis of mutual influence in collectives.

INTRODUCTION

Coherent collective behavior fascinates us when a global pattern emerges from individuals who share information with others only in their local vicinity. Decisions made by one individual apparently cascade throughout the entire group. That said, not all members have the same influence. The challenge here in explaining these emergent behaviors is to infer the underlying relationships among individuals from observations. Expectedly, this challenge has attracted many over decades to diagnose collective behaviors in a variety of systems (*1*–*7*). In epithelial Madin-Darby canine kidney monolayers, for example, collective cell migration is triggered by multicellular protrusions, which form fingerlike structures (*8*–*10*). Photoablating a group of cells from the fingertip makes the remaining cell group lose its sense of direction. The interpretation is that the former and latter cells have acted as if they were "leaders" and "followers." This functional assignment can be carried out because of their spatial location along the protrusion and because the two cell classes are genetically distinct (*10*).

Understanding the relationships between leaders and followers—even defining what those roles mean (*11*)—is very difficult, especially so, when probing the mechanisms that cause the dynamical behaviors of aggregated agents. Beyond cells, these basic questions also apply to bird flocking (*3*), fish schooling (*4*, *5*), caribou migration (*6*), and baboon foraging (*7*). A leader agent is often defined as an individual that influences others more than others influence it. That is, the role is fundamentally asymmetric. Previous studies (*4*, *5*, *12*–*16*) proposed that, under this definition, pairwise analysis of trajectories can assign leaders and followers under the working hypothesis that a change in motion of the leader forecasts a change in motion of the follower. From this, one interprets the change of leader motion as a candidate cause that triggers the motion of followers.

Various statistical quantities are used to infer causal relationships (*17*). In pigeon flocks, for example, time-delayed correlation between the orientation of individuals at one time instance and the orientation of others at previous times reveals a hierarchical leadership structure and also provides a method to quantify the time scale of influence (*3*). In such a case, the motion of one pigeon is correlated with the past motion of another. Granger causality (*18*) is seen as an improvement to time-delayed correlation as it quantifies the predictability of the current state of a variable based on knowledge of a variable at a previous time. Time-delayed correlation and Granger causality both assume linear relationships between variables, though. This generally does not hold. More recent studies argued that information-theoretic quantities—transfer entropy (TE), time-delayed mutual information (TDMI), and causation entropy—are superior when quantifying influence since they naturally accommodate the highly nonlinear nature of multiagent systems (*5*, *12*–*14*, *19*–*31*).

In practice, one must consider the potential for misclassifying influence when using information-theoretic methods. Carefully considering the definitions of information-theoretic quantities—such as TE or TDMI—further illuminates the types of influence a particular individual has. As pointed out by Schreiber (*32*) in introducing TE, TDMI reports a nonzero value between the present of a stochastic variable X and the future of a stochastic variable Y even when X has no direct influence on Y. This implies that it cannot be directly used to infer the underlying mutual influence among individuals. It also includes additional information not intrinsically coming from X.

TE from X to Y, in contrast, computes the reduction of uncertainty about Y's future while knowing X's present, conditioned on Y's present. Recently though, the study of James *et al.* (*33*) showed that, paralleling TDMI, TE incorporates additional, unwarranted information, namely, the reduction of uncertainty about Y that occurs by knowing the present state of X and Y simultaneously. This information is extraneous to determining "flow" and, misleadingly, adds to the desired information: intrinsic flow from X to Y. In this view, TE decomposes into two distinct modes of information flow—intrinsic and synergistic (*33*).

In addition to the above drawbacks, any measure of causality or information flow suffers from the problem of hidden variables, for example, when an outsider influences two agents concurrently. By

[1]Research Center of Mathematics for Social Creativity, Research Institute for Electronic Science, Hokkaido University Kita 20, Nishi 10, Kita-ku, Sapporo, Hokkaido 001-0020, Japan. [2]Pabna University of Science and Technology, Pabna 6600, Bangladesh. [3]Reddit Inc., 420 Taylor Street, San Francisco, CA 94102, USA. [4]Department of Physics, Complexity Sciences Center, University of California, Davis, Davis, CA 95616, USA. [5]École Normale Supérieure de Rennes, Robert Schumann, Campus de, Av. de Ker Lann, 35170 Bruz, France. [6]Institute for Chemical Reaction Design and Discovery (WPI-ICReDD), Hokkaido University Kita 21 Nishi 10, Kita-ku, Sapporo, Hokkaido 001-0021, Japan. [7]Graduate School of Chemical Sciences and Engineering Materials Chemistry and Energy Course, Hokkaido University Kita 13, Nishi 8, Kita-ku Sapporo, Hokkaido 060-0812, Japan.
*Corresponding author. Email: tamiki@es.hokudai.ac.jp

Downloaded from https://www.science.org on March 02, 2022

observing the two agents alone, one may infer a causal relationship between them. One can condition on other agents by computing causation entropy, which is an extension of TE that accounts for the effects of additional agents (*28*). The difficulty with this approach is that the dimension of the probability distribution required for computing the measure grows exponentially with respect to the number of additional variables, which may require unfeasibly large amounts of data to properly sample.

Pairwise interactions are fundamental to information theory's development of input-output ("two-port") communication channels (*34*). Hence, they provide a primary statistical tool that, as we show, makes it possible to infer the underlying influences among individuals. To obtain maximum insight into the mechanisms underlying multiagent systems, the following focuses on decomposing TE and TDMI into James *et al.*'s (*33*) three different fundamental modes of information flow—termed intrinsic, shared, and synergistic information flows. The results demonstrate how the decomposed elemental information flows shed light on the influences that drive leader-follower relationships. This work augments the previous work in (*33*) by manifesting previously unidentified features that come out from the mode decomposition of most commonly used information measures, such as extracting the nature of multiagent interactions from only pairwise trajectories and shedding light on previously unidentified interpretations of the modes of information flow that provide us with a more solid interpretation of causal inference than TE.

As an illustrative vehicle, we use a generalized Vicsek model (*35*) with two additional features: (i) tunable influence weight of one particle over another (i.e., leaders have larger influence) and (ii) particle memory. We show that, by analyzing the effects of (i) and (ii) on the three modes of information flow, intrinsic information flow exists whenever the motion of an agent depends on another with nonzero weight, as does TE and TDMI. Shared and synergistic information, however, can or cannot occur depending on the setting of (i) and (ii), as we will show further. These results extend previous studies on modes of information flow in finite-state hidden Markov models by introducing Vicsek models that are fundamental to understanding collective motion. In addition, the results manifest that decomposing pairwise information flow between two agents (e.g., leader and follower) enables us to differentiate the underlying interaction pattern driven by some additional agents (e.g., followers) in more than two interacting agents and also allows us to probe the effect of agent memory on the different modes of information flow and their role in collective behavior.

Background: Measuring causal influence

We now review several information-theoretic measures of statistical interdependence—measures that have been offered up as ways to detect causal influence. With these in hand, we turn to explore how useful (or not) they are in analyzing leader-follower relationships.

Detecting causal influence via TDMI and TE

Our definition says that leaders are, on average, more influential than followers. As a consequence of this asymmetry, a follower's behavior is affected by the leaders, but there is a time delay. To determine the degree of causal influence between random variables, quantitative measures have been introduced from information theory, such as TDMI and transfer (conditional) entropy. As they make no assumption about the functional relationship between variables, the latter improve on the more commonly used measures of time-delayed correlation (*13*) and Granger causality (*18*), which can capture only linear functional relationships.

Consider two stationary stochastic processes $X = (\ldots, x_{t-1}, x_t, x_{t+1}, \ldots)$ and $Y = (\ldots, y_{t-1}, y_t, y_{t+1}, \ldots)$ with probability mass functions $p(x_t) = Pr\{X = x_t\}$ and $p(y_t) = Pr\{Y = y_t\}$, respectively. TDMI from X to Y with time delay τ is given by (*27*)

$$\mathcal{M}_{X\to Y}(\tau) = I(X_t; Y_{t+\tau}) = H(Y_{t+\tau}) - H(Y_{t+\tau} \mid X_t) = H(X_t) - H(X_t \mid Y_{t+\tau}) \quad (1)$$

where $H(Y_{t+\tau})$ and $H(Y_{t+\tau} \mid X_t)$ are the Shannon entropy and conditional entropy, respectively. They measure, in turn, the uncertainty in $Y_{t+\tau}$ and the uncertainty in $Y_{t+\tau}$ remaining given X_t, respectively. In other words, being their difference, the mutual information $\mathcal{M}_{X\to Y}(\tau)$ monitors the reduction in uncertainty in Y's future knowing X's at a time t. Since mutual information is symmetric, this is also the reduction of uncertainty in X's present knowing Y at the future time $t + \tau$. The symmetry, though, means that it cannot be used to infer causal influence since, by assumption, the future cannot influence present.

In addition, TDMI has another perhaps more subtle drawback: When predicting influence, it can be nonzero when two variables have shared history (*32*). That is, the condition $\mathcal{M}_{X\to Y}(\tau)$ may hold when variable Y is not directly influenced by variable X but when either X or Y dynamics contains memory of their present configurations.

TE was introduced to overcome these shortcomings (*32*): If X influences Y, then predicting Y's future becomes easier after knowing the present of both X and Y, compared to only knowing Y's present. The TE from X to Y takes the form of conditional mutual information

$$\mathcal{T}_{X\to Y}(\tau) = I(X_t; Y_{t+\tau} \mid Y_t) = H(Y_{t+\tau} \mid Y_t) - H(Y_{t+\tau} \mid Y_t, X_t) \quad (2)$$

That is, $\mathcal{T}_{X\to Y}(\tau)$ is TDMI between Y at time $t + \tau$ and X at time t conditioned by Y at time t. It is the same as subtracting the uncertainty remaining in Y at time $t + \tau$ given both X and Y at the present time t from that in $Y_{t+\tau}$ given Y_t. The latter corresponds to the uncertainty of $Y_{t+\tau}$ reduced by knowing X_t in addition to the knowledge of Y_t.

$\mathcal{T}_{X\to Y}$ has become one of the standard methods for measuring statistical influence in classifying leaders and followers (*4*, *5*, *13*–*15*, *25*, *26*). More broadly, since it improves upon TDMI for quantifying asymmetric relationships, it has become a standard for inferring causal relationships in many areas of science, including neuroscience (*23*, *36*–*40*), chemistry (*41*), human behavior (*42*, *43*), and Earth systems (*44*–*46*).

We note, however, that, like correlation, information-theoretic quantities such as TDMI and TE are not sufficient in themselves to identify causality. The latter also requires accounting for the influence of latent or hidden variables. TDMI does not condition on the present or past time steps, and conditioning on the history in computing TE is finite in time length (Y_t in Eq. 2). Therefore, each variable's history may act like a hidden variable that influences outcomes. This, in turn, can lead to spurious effects when estimating information flow, as we will elucidate shortly.

Recently, it was demonstrated for a simple binary system that $\mathcal{T}_{X\to Y}(\tau) > 0$ can occur although knowledge of X_t alone cannot reduce the uncertainty in $Y_{t+\tau}$ (*33*). It was pointed out that, in addition to information intrinsic to reducing uncertainty in $Y_{t+\tau}$, which comes from knowing X_t, independent of Y_t, TE from X to Y includes

Downloaded from https://www.science.org on March 02, 2022

information that reduces the uncertainty in $Y_{t+\tau}$, which comes from knowing X_t and Y_t simultaneously (*47*).

Here, intrinsic information flow is the additional reduction in uncertainty in Y's future gained from knowing X's present compared to the reduction from knowing Y's present alone or knowing the present of X and Y simultaneously. Intrinsic mutual information (IMI) was proposed (*47*, *48*) as a measure that avoids including influence that comes from both the present of X and Y when predicting Y's future. Note that, while IMI is a specific quantity and not synonymous with intrinsic information flow, it can be seen as an attempt to compute intrinsic information flow between two variables.

Diagnosing causal influence via intrinsic, shared, and synergistic information

Intrinsic information flow from X to Y is regarded as the information flowing from the present of X (X_t) to the future of Y ($Y_{t+\tau}$), which cannot be attributed in any way to the present of Y (Y_t), in contrast to TE $\mathcal{T}_{X\to Y}$ attributed in part to Y_t. Independently, secret key agreement rate $S(A;\, B\,|\,C)$ is a key concept of cryptography, which quantifies the rate of secret (information) shared between two stochastic variables A (say, Alice) and B (Bob) free from a third variable C (Eve), regarded as an "eavesdropper" (see section S3 and fig. S15) (*48*, *49*). By using the secret key agreement rate where A, B, and C are replaced by X_t, $Y_{t+\tau}$, and Y_t, the amount of information flowing intrinsically from X to Y is equal to the secret key agreement rate $S(X_t;\, Y_{t+\tau}\,|\,Y_t)$ (defined in section S3) (*33*). Since $S(X_t;\, Y_{t+\tau}\,|\,Y_t)$ is not computable in practice, IMI $\mathcal{I}_{X\to Y}$ is used as a (workable) upper bound on $S(X_t;\, Y_{t+\tau}\,|\,Y_t)$ to monitor information that flows intrinsically from X to Y.

The amount of IMI communicated from process X to another Y is the infimum of $I(X_t;\, Y_{t+\tau}\,|\,\overline{Y}_t)$ taken over all possible conditional distributions $p(\overline{Y}_t\,|\,Y_t)$ (*33*)

$$\mathcal{I}_{X\to Y}(\tau) := \inf\left\{ I(X_t;\, Y_{t+\tau}|\,\overline{Y}_t) : p(X_t, Y_{t+\tau}, \overline{Y}_t) = \sum_{y\in Y_t} p(X_t, Y_{t+\tau}, Y_t = y)\, p(\overline{Y}_t | Y_t = y) \right\} \tag{3}$$

Here, $\overline{Y}_t$ is an auxiliary variable used to realize the upper bound of $S(X_t;\, Y_{t+\tau}\,|\,Y_t)$. It satisfies the Markov property $Y_t \to \overline{Y}_t$, and $\overline{Y}_t$ has no information to be shared with X_t and $Y_{t+\tau}$, conditioned by Y_t, i..e, $I(X_t\, Y_{t+\tau};\, \overline{Y}_t\,|\,Y_t) = 0$. Here, $A \to B$ signifies that B depends only on A, and the infimum is taken over all possible conditional distributions $p(\overline{Y}_t\,|\,Y_t)$.

$\mathcal{I}_{X\to Y}$ represents uncertainty reduction in Y's future that comes from knowing only X's present as much as possible under the assumption of the Markov property with respect to Y and $\overline{Y}$, which has been found to be a convenient and accurate bound on the secret key agreement rate $S(X_t;\, Y_{t+\tau}\,|\,Y_t)(\le \mathcal{I}_{X\to Y})$ (*47*, *48*).

The intuitive explanation as to why the IMI forms an upper bound on the secret key agreement rate is as follows: Alice (A) and Bob (B) cannot be said to hold any information secretly if they are all accessible to Eve (C), including their secret key to decrypt their public communication (see also section S3). How does Eve infer what information Alice and Bob exchange solely from the knowledge of communication between Alice and Bob? Suppose that Eve is not restricted to using solely the value c of the stochastic variable C she observes at each step but rather any possible statistical transformation of the variable C represented by $C' \sim p(C'\,|\,C)p(C)$ with an auxiliary variable C' [Note $p(C'\,|\,C) = \delta_{C,C'}$ (δ:Kronecker delta) corresponds to using the original C in the inference]. Said another way, Eve has access to knowledge regarding A and B via C plus any manipulation or transformation of C, and so the secret key agreement rate between Alice and Bob is certainly bound from above by conditional mutual information $I(A;B\,|\,C'_M)$, where C'_M is the transformation of C that minimizes the mutual information between A and B.

How does this intuition translate to the information flow setting? Intrinsic information flow from X to Y is a very restricted concept, that is, it is the information flowing from the present of X (X_t) to the future of Y ($Y_{t+\tau}$), which is free from the present of Y (Y_t). So in a very similar sense as above, if there is some manipulation of Y_t (i.e., $\overline{Y}_t$), which reduces the information between X_t and $Y_{t+\tau}$, $I(X_t;\, Y_{t+\tau}\,|\,\overline{Y}_t)$ is expected not to be attributed to the present of Y (since it was inferred via the present of Y). It is in this sense that using the secret key agreement rate helps us quantify intrinsic information flow.

The following relations delineate the importance of $\mathcal{I}_{X\to Y}(\tau)$ and its relationship to $S(X_t;\, Y_{t+\tau}\,|\,Y_t)$, $\mathcal{T}_{X\to Y}(\tau)$, and $\mathcal{M}_{X\to Y}(\tau)$

$$0 \le S(X_t;\, Y_{t+\tau}\,|\,Y_t) \le \mathcal{I}_{X\to Y}(\tau) \tag{4}$$

$$\mathcal{I}_{X\to Y}(\tau) \le \mathcal{T}_{X\to Y}(\tau) \le H(Y_{t+\tau}\,|\,Y_t) \tag{5}$$

and

$$\mathcal{I}_{X\to Y}(\tau) \le \mathcal{M}_{X\to Y}(\tau) \tag{6}$$

In effect, Eqs. 4 to 6 demonstrate that IMI can be used to compute bounds on the deviations of $\mathcal{M}_{X\to Y}(\tau)$ and $\mathcal{T}_{X\to Y}(\tau)$ from $S(X_t;\, Y_{t+\tau}\,|\,Y_t)$. That is, whenever equality does not hold in Eq. 5 (Eq. 6), then there must be a portion of $\mathcal{T}_{X\to Y}(\tau)$ ($\mathcal{M}_{X\to Y}(\tau)$) that is not intrinsically coming from X. From here on, we set $\tau = 1$ and omit τ from equations, as it has been shown that $\tau = 1$ best captures the information flow between two particles in the Vicsek model (*12*).

Once $\mathcal{I}_{X\to Y}$ has been determined, shared $\sigma_{X\to Y}$ and synergistic $\mathcal{S}_{X\to Y}$ information follow immediately by subtracting it from TDMI ($\mathcal{M}_{X\to Y}$) and TE ($\mathcal{T}_{X\to Y}$), respectively

$$\sigma_{X\to Y} = \mathcal{M}_{X\to Y} - \mathcal{I}_{X\to Y} \tag{7}$$

and

$$\mathcal{S}_{X\to Y} = \mathcal{T}_{X\to Y} - \mathcal{I}_{X\to Y} \tag{8}$$

From Eqs. 4 to 6, one sees that

$$0 \le \sigma_{X\to Y} \le \mathcal{M}_{X\to Y} \tag{9}$$

and

$$0 \le \mathcal{S}_{X\to Y} \le \mathcal{T}_{X\to Y} \tag{10}$$

Note that $\mathcal{I}_{X\to Y}$ monitors information coming from (mostly) X alone to Y, since IMI provides an upper bound on $S(X_t;\, Y_{t+1}\,|\,Y_t)$. Because of this, $\mathcal{S}_{X\to Y} > 0$ ($\sigma_{X\to Y} > 0$) implies that $\mathcal{T}_{X\to Y}$ ($\mathcal{M}_{X\to Y}$) contains information that comes from Y's present and that $\mathcal{T}_{X\to Y}$ ($\mathcal{M}_{X\to Y}$) should not be interpreted as information flowing only from X to Y in those cases.

In the following, on the basis of $\mathcal{I}_{X\to Y}$ being (mostly) information coming from only X to Y, $\sigma_{X\to Y}$ is the part of $\mathcal{M}_{X\to Y}$, which comes from knowing both variables. On postulating that it is the information redundant in both X and Y, we referred to it as shared

Downloaded from https://www.science.org on March 02, 2022

information. Similarly, $\mathcal{S}_{X\to Y}$ is synergistic information since it is the part of $\mathcal{T}_{X\to Y}$, which comes from knowing both variables and since it arises from simultaneously knowing X and Y at present (*33*). Together, $\mathcal{I}$, σ, and $\mathcal{S}$ reveal a much more detailed decomposition of the relationship between X and Y than can be inferred from $\mathcal{T}$ or $\mathcal{M}$ alone.

To ground this information-theoretic setting, the following shows, using a modified Vicsek model of collective behavior, that $\mathcal{T}$ and $\mathcal{M}$ can result in misleading interpretation concerning the underlying actual relationship among individuals. We go on to propose that $\mathcal{I}$, σ, and $\mathcal{S}$ provide a firmer interpretation of the relationship without requiring additional experiments.

RESULTS

With this background, we now turn to diagnose interactions between individuals in a collective system and between individuals and the collective.

Modified Vicsek model

To demonstrate the interpretability of informational modes $\mathcal{I}$, σ, and $\mathcal{S}$, we introduce a series of augmented Vicsek models. These extend the original (*35*) with asymmetric interactions and turn on-off the dependence on the present dynamics of interacting particles.

Consider N particles lying within a square box of length L with periodic boundary conditions. Particle i's position $\vec{\mathbf{r}}_i^t$ at time t is updated over time increment Δt according to

$$\vec{\mathbf{r}}_i^{t+1} = \vec{\mathbf{r}}_i^t + \vec{\mathbf{v}}_i^t \Delta t \tag{11}$$

where $\vec{\mathbf{v}}_i^t$ denotes particle i's velocity at time t and $i = 1,2, \ldots, N$. For simplicity, particles have uniform constant speed v_0, and only their orientations θ_i change.

Particle orientation is updated at each time by taking the weighted average of the velocity of neighboring particles within a given radius R

$$\theta_i(t+1) = \langle \boldsymbol{\theta}(t) \rangle_{R,\mathbf{w},\vec{\mathbf{r}}_i^t} + \Delta\theta_i \tag{12}$$

where $\mathbf{w}$ is a nonnegative asymmetric matrix whose w_{ij} element determines the interaction strength that particle i exerts on particle j. $w_{ij} > w_{ji}$ whenever particle i is a leader, and particle j is a follower in our setting. To model thermal noise, $\Delta\theta_i$ is a random number uniformly distributed in the range $[-\eta_0/2, \eta_0/2]$ and is chosen uniquely for each particle i at each time step. In the original model (*35*), the right-hand side of Eq. 12 ensured that $\theta_i(t+1)$ resulted from the configurations of all the particles j (including that of the same particle i) within the circle of radius R centered at $\vec{\mathbf{r}}_i^t$.

Now, consider modified dynamics that modulate the dependence on $\theta_j(t)$ associated with follower-leader interactions that determine $\theta_i(t+1)$: The leader influences the follower, but the follower does not influence the leader; i.e., $w_{\mathrm{LF}} > 0$, while $w_{\mathrm{FL}} = 0$.

To graphically understand our models, we show Fig. 1 (Aa to Da), where each graph depicts one possible interaction protocol in a simple, two-particle system that determine $\theta_i(t+1)$, where L and F denote leader and follower, respectively. There, if A or B are either L or F, then $A \to B$ signifies that A's present state influences the B's future state. We vary $w_{\mathrm{LF}} \in [1,10]$. We set $w_{\mathrm{LL}} = 1$ and $w_{\mathrm{FF}} = 1$ for the models in which the present state of L (F) influences the future state of L (F). For models in which the present does not influence the future for the same particle (F or L)—see Fig. 1Aa (i.e., L's and F's dynamics), Ba (L's), and Ca (F's)—we replace $\theta_i(t)$, which appears in computing $\langle \boldsymbol{\theta}(t) \rangle_{R,\mathbf{w},\vec{\mathbf{r}}_i^t}$ (see Σ′ term in Eq. 20) by a random number in the interval $[0,2\pi]$ to erase any influence from $\theta_i(t)$'s present. Note that the value of w_{LL} (w_{FF}) is inconsequential when the dynamics of leader (follower) do not depend on their present state, and $\theta_i(t)$ depends solely on a random number in the interval $[0,2\pi]$. In type A, neither the future dynamics of L nor F depend on their present (Fig. 1, Aa). In type B, only the future dynamics of F depends on its present (Fig. 1, Ba). In type C, only the future dynamics of L depends on its present (Fig. 1, Ca). In type D, both L's and F's future dynamics depend on their present (Fig. 1, Da). To further examine the effects of the history of L in types C and D, we also introduced interaction types C′ and D′, which are the same as interaction types C and D, respectively, except the future state of L depends on its present only when time step t is even, and L "forgets" its present in its future dynamics as in types A and B whenever t is odd. In types C′ and D′, the dependence of the future of F on its present are not changed, that is, the future of F does not depend on its present in type C′, and the future of F always depends on its present in type D′ regardless of the value of t.

Informational modes between leaders and followers

Misinterpreting causal influence

Let us first examine the amounts of TDMI ($\mathcal{M}$) and TE ($\mathcal{T}$) shown in Fig. 1 for different interaction types, integrated over ranges of both w_{LF} and η_0 (the landscapes of $\mathcal{M}$ and $\mathcal{T}$ as a function of w_{LF} and η_0 are shown in figs. S1 and S2). Here, the white circle with diagonal shading represents $\mathcal{M}$, and the red circle represents $\mathcal{T}$. The Venn diagrams can graphically capture the relationship among the decomposed modes of information with TDMI and TE. The overlapping region between $\mathcal{M}$ and $\mathcal{T}$, the part of $\mathcal{M}$ that is not overlapping with $\mathcal{T}$, and the part of $\mathcal{T}$ that is not overlapping with $\mathcal{M}$ represent $\mathcal{I}$, σ, and $\mathcal{S}$, respectively, and are discussed in the following section.

As expected, $\mathcal{M}_{\mathrm{L}\to\mathrm{F}}$ and $\mathcal{T}_{\mathrm{L}\to\mathrm{F}}$ in Fig. 1(Ab to Db and Cd to Dd) exhibit substantial, nonzero values in all interaction types, since L is influencing F in all cases. Naively, one expects $\mathcal{M}_{\mathrm{F}\to\mathrm{L}}$ and $\mathcal{T}_{\mathrm{F}\to\mathrm{L}}$ to be zero for all cases since F does not influence L at all. However, there are spurious values of both $\mathcal{M}_{\mathrm{F}\to\mathrm{L}}$ and $\mathcal{T}_{\mathrm{F}\to\mathrm{L}}$ in Fig. 1(Cc to Dc), as well as spurious values of $\mathcal{T}_{\mathrm{F}\to\mathrm{L}}$ alone in Fig. 1(Ce to De). Spurious amounts of $\mathcal{T}_{\mathrm{X}\to\mathrm{Y}}$ and $\mathcal{M}_{\mathrm{X}\to\mathrm{Y}}$, even in cases where X does not influence Y, have been ignored by a large and growing body of research in quantifying causal relationships. The following section elaborates on how decomposing $\mathcal{M}$ and $\mathcal{T}$ into $\mathcal{I}$, σ, and $\mathcal{S}$ can improve the interpretation of information flow using different interaction types as examples.

Modes of information flow

We now interpret by Venn diagrams the different modes of information flow integrated over ranges of w_{LF} and η_0 for each interaction type (Venn diagrams for analogous binary systems are shown in fig. S13 and described in section S3). The landscapes of $\mathcal{I}$, σ, and S as a function of w_{LF} and η_0 are shown in figs. S3 to S5 for types A, B, C, and D and in figs. S6 and S7, respectively, for types C′ and D′. Since each interaction type corresponds to a special case of information flow, let us briefly examine each one.

In type A, where neither L nor F depends on its present (Fig. 1, Aa), only $\mathcal{I}_{\mathrm{L}\to\mathrm{F}}$ and no other type of information flow is observed, as seen in Fig. 1 (Ab). In this case, since $\mathcal{M}$ and $\mathcal{T}$ overlap completely,

Downloaded from https://www.science.org on March 02, 2022

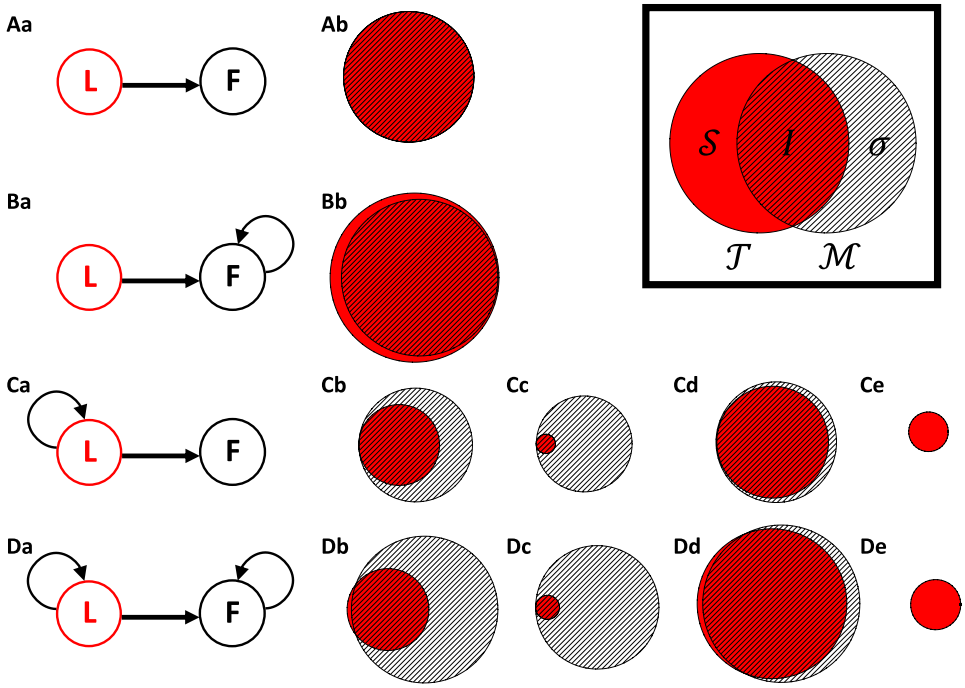

Fig. 1. Graph representation of interaction types A, B, C, and D and the corresponding Venn diagrams representing information flow. Areas of the circles are computed by integrating $\mathcal{M}$ and $\mathcal{T}$ over η_0 ranging from 0 to 2π and w_{LF} ranging from 1.0 to 10.0. The area of red circles and white striped circles in the Venn diagrams are equal to the (integrated) TDMI $\int\mathcal{M}(\eta_0, w_{LF})\, d\eta_0\, dw_{LF}$ and TE $\int\mathcal{T}(\eta_0, w_{LF})\, d\eta_0\, dw_{LF}$, respectively. The centers of the two circles are determined as follows: First, each of the centers is connected with a horizontal line without being overlapped [the TE (TDMI) is located at the left (right)], and then a binary search algorithm was used to find the placement of those circles whose overlapping area is equal to the IMI $\int\mathcal{I}(\eta_0, w_{LF})\, d\eta_0\, dw_{LF}$ by decreasing the distance between the centers. The part of the red (white striped) circle not overlapped with the white striped (red) circle has area equal to the synergistic information $\int\mathcal{S}(\eta_0, w_{LF})\, d\eta_0\, dw_{LF}$ (the shared information $\int\sigma(\eta_0, w_{LF})\, d\eta_0\, dw_{LF}$) (see Legend). (**Aa**) Type A. (**Ba**) Type B. (**Ca**) Type C. (**Da**) Type D. (**Ab** to **Db**) Venn diagrams from leader to follower for interaction types A to D. The information flows from follower to leader for types A and B are negligible and therefore are not shown. (**Cc** to **Dc**) Those from follower to leader for interaction types C and D. (**Cd** to **Dd**) Those from leader to follower for interaction types C′ and D′. (**Ce** to **De**) Those from follower to leader for interaction types C′ and D′.

there is no nonoverlapping region of either $\mathcal{M}$ or $\mathcal{T}$, signifying that $\mathcal{I} = \mathcal{M} = \mathcal{T}$ and $\sigma = \mathcal{S} = 0$. Furthermore, all F to L information flow quantities are equal to zero, thus the Venn diagram for type A from F to L is not shown. Although this may not be a realistic case in systems of interacting agents, since agents are likely to depend not only on each other but also their history, it is only the case demonstrated where either $\mathcal{T}$ or $\mathcal{M}$ can accurately convey causal relationships between agents.

Figure 1 (Ba) represents type B where only F depends on its present in the $\theta_i(t + 1)$ dynamics. $\mathcal{I}_{L\to F}$ and $\mathcal{S}_{L\to F}$, (and thus $\mathcal{M}_{L\to F}$ and $\mathcal{T}_{L\to F}$) (Fig. 1, Bb) all increase compared to the case where F's dynamics do not depend on its present (Fig. 1, Ab). This further emphasizes our point that dependence on the present state plays a key role in the calculation of information flow even when the interactions between individuals are not intrinsically changing. Notably, the red sliver of nonoverlapping region between $\mathcal{M}$ and $\mathcal{T}$ in Fig. 1 (Bb) shows the appearance of synergistic information $\mathcal{S}_{L\to F}$, which denotes that part of $\mathcal{T}_{L\to F}$ is not intrinsically coming from L. Since F depends on its present, the simultaneous knowledge of the present state of F and L provides more predictive power than knowing either the present state of F or L alone. Also to be noted is that $\sigma_{L\to F}$ is equal to zero, which is due to L not imparting any of its history onto F. F to L information flows in this case are again equal to zero, since the leader has no memory of its past history to share with the follower.

Figure 1 (Ca) represents type C, where only L depends on its present. In this case, a substantial amount of $\sigma_{L\to F}$ appears because of the dependence of the future state of both L and F on the present state of L, as shown in Fig. 1 (Cb). L imparts information from its present on to the future dynamics of F, and meanwhile, this information is already contained in the future state of F because of the dependence of the future state of F on that same history (i.e., L's present). In contrast to types A and B, there is a substantial amount of shared history between F and L, as shown in Fig. 1 (Cb). As in the binary system proposed by Kaiser and Schreiber (*50*) having the same graph representation as Fig. 1 (Ca), there exists a substantial amount of $\mathcal{M}_{F\to L}$ although there is no direct interaction in that direction. By decomposing $\mathcal{M}_{F\to L}$ into $\mathcal{I}_{F\to L}$ and $\sigma_{F\to L}$, we quantitatively show that the spurious amount of $\mathcal{M}_{F\to L}$ is coming solely from $\sigma_{L\to F}$ and is thus not intrinsically coming from F. TE $\mathcal{T}$ was introduced to reconcile this issue (*32*). $\mathcal{T}_{F\to L}$ does in fact reduce the amount information flow in that direction in our model, given that $\mathcal{T}_{F\to L}$ is notably less than $\mathcal{M}_{F\to L}$ in Fig. 1 (Cc).

Why then, are $\mathcal{T}_{F\to L}$, and more importantly, $\mathcal{I}_{F\to L}$ not equal to zero in type C (Fig. 1, Cc)? Surely, information is not intrinsically flowing from F to L in this case since there is no direct link from F to L in Fig. 1 (Ca). The reason is that the history of L acts as a hidden variable, imparting information onto both L and F. To verify this, we have introduced interaction type C′, which is the same as interaction type C (Fig. 1, Ca), except that the future state of L depends

Downloaded from https://www.science.org on March 02, 2022

on its present only when time step t is even, and L forgets its present in its future dynamics when time step t is odd. Although the present state of L can still act as a hidden variable imparting information on both L and F, the computation of $\mathcal{I}_{F\to L}$ conditions on the present state of L and also minimizes the uncertainty coming from the present state of L as much as possible. Therefore, it is not possible for the history of L to have any influence on the value of $\mathcal{I}_{F\to L}$ in type C′ since it remembers at most one time step of its past history, which is being minimized in the computation of $\mathcal{I}$. The Venn diagram from follower to leader in this case is shown in Fig. 1 (Ce). Note that $\mathcal{T}_{F\to L}$ still exists in this case but only in the form of synergistic information $\mathcal{S}_{F\to L}$ as $\mathcal{I}_{F\to L} = 0$. Recall that the existence of $\mathcal{S}_{F\to L}$, which is a part of TE, implies that simultaneous knowledge of the present states of L and F allows for an improvement on the prediction of the future of L compared to individual knowledge of the present states of L or F alone. Why does this happen under $\mathcal{I}_{F\to L} = 0$? Here, we explain the origin of the existence of synergistic information intuitively. For interaction type C′, $\theta_F(t + 1)$ always results from $\theta_L(t)$ irrespective of the time step. In turn, the present configuration of $\theta_L(t)$ affects its future at every even time step t. However, $\theta_L(t + 1)$ is taken from $[0,2\pi)$ randomly to reset its history every odd time step t, i.e., $\theta_F(t + 1) \neq \theta_L(t + 1)$. That is, simultaneous knowledge of the states of L and F reduces uncertainty about whether time step is even or odd, and, therefore, simultaneous knowledge of the present states of L and F improves the prediction power of the future of L compared to solely individual knowledge of the present state of L and F (see more in detail in section S2).

Last, Fig. 1 (Da) represents type D, which is perhaps the most intuitive case in typical systems, where both entities depend on their present state in the $\theta_i(t + 1)$ dynamics. $\mathcal{M}$, $\mathcal{T}$, and σ are greater than zero for similar reasons as they are in types B and C for both leader to follower and follower to leader, and $\mathcal{S}_{L\to F}$ but not $\mathcal{S}_{F\to L}$ is greater than zero for similar reasons as well. Type D′ is the same as C′, except F depends on its present dynamics as in type D. As in type C′, the only type of information flow from F to L in type D′ is $\mathcal{S}_{F\to L}$, and for similar reasons. Thus, type D, representing the most typical types of multiagent systems, contains a rich profile of information flows, which we have explained by analyzing types A, B, C, C′, and D′.

Systems of more than two agents

The analysis up to now addressed only pairwise interactions. This is in accord with the theoretical basis of the information measures used; for example, $\mathcal{T}_{X\to Y}(\tau)$ in Eq. 2. The measures generalize straightforwardly to account for additional time series, say, of a third particle (or agent); see, for example, the causation entropy (*26*). Suppose the third variable Z, in addition to X and Y, are each symbolized by m discrete values. Then, for example, the dimension of the probability distribution $p(Y_{t+1}, X_t, Z_t)$ is $m^3 - 1$ (−1 is because of probability normalization). This means that, the more the number of additional variables to be conditioned on increases, the more the dimension of the probability distribution required for computing the measures grows exponentially with respect to the number of additional variables. This requires increasingly large amounts of data to properly sample. Therefore, in multiagent systems, it is not usually feasible to condition on all or even a few other agents that interact with a given agent. In addition, even if additional variable(s) that indirectly affect(s) interactions between X and Y exist, it is nontrivial to look for this indirect "cause." These hidden variables may be another agent entity, some past memory of the process of X and/or Y longer than being taken into account in the elucidation of TE, or something else.

Nonetheless, estimating two-agent information measures has been proven useful for monitoring influence in systems having more than two agents (*5*, *13*, *14*, *25*, *27*). We will now show how measuring $\mathcal{I}$, σ, and $\mathcal{S}$ gives marked improvements even in these admittedly approximate settings.

Consider a collective in which L and F mutually interact with one another, but under model A, followers also directly interact with each other, and under model B, they do not. See, for example, Fig. 2 for the case of three agents. In the following discussion, there is one leader agent, and the number of follower agents N_F is varied. L refers to the leader, and F refers to a particular follower.

Figure 3 (A and B) displays $\mathcal{M}_{L\to F}$ for models A and B, respectively. The plots of σ (see fig. S9) are almost indistinguishable from those of $\mathcal{M}$, indicating that a majority of $\mathcal{M}$ is actually coming from σ, which is due to shared history between L and F. As has been established for the Vicsek model (*35*), cohesive behavior increases as a function of density. Here, $\mathcal{M}_{L\to F}$ and $\sigma_{L\to F}$ increase as a function of N_F in model A. Model B, however, is not the same as the original Vicsek model in that followers do not interact with each other, and therefore, $\mathcal{M}_{L\to F}$ and $\sigma_{L\to F}$ decrease as a function of N_F, since the inclusion of additional agents that are not interacting decreases the overall cohesion between the present state of L and the future state of F. The plots of $\mathcal{M}_{F\to L}$ for models A and B are not shown as they are not distinguishable by eye from those of $\mathcal{M}_{L\to F}$ (see fig. S8).

Figure 4 (A to D) shows $\mathcal{T}_{L\to F}$ as a function of η_0 for model A, $\mathcal{T}_{F\to L}$ for model A, $\mathcal{T}_{L\to F}$ for model B, and $\mathcal{T}_{F\to L}$ for model B, respectively. At $\eta_0 = 0$, agent movements quickly reach a regular parallel flow independent of initial coordinates and velocities, and thus, any information about their present orientations are negligible (on average) in predicting the others' orientational motions (see movie S1). In practice, all agents are subject to finite noise due to their environment (represented here by thermal fluctuation). Gradual decreases of $\mathcal{T}$ as η_0 increases simply arise from this natural stochasticity. In both models A and B, there are small bumps in $\mathcal{T}_{L\to F}$ and $\mathcal{T}_{F\to L}$ at $\eta_0 \simeq 0.7\pi$, but the notable difference is that the bumps clearly decrease as a function of N_F from L to F and F to L in model A (Fig. 4, A and B) and from F to L in model B (Fig. 4D) but not from L to F in model B (Fig. 4C).

The existence of bumps in $\mathcal{T}$ at $\eta_0 \simeq 0.7\pi$ and the difference in their behavior between models A and B can be explained by decomposing $\mathcal{T}$ into $\mathcal{I}$ and $\mathcal{S}$. Figure 5 (A to D) shows $\mathcal{I}_{L\to F}$ as a function of η_0 for model A, $\mathcal{I}_{F\to L}$ for model A, $\mathcal{I}_{L\to F}$ for model B and $\mathcal{I}_{F\to L}$ for model B, respectively. While the overall trend is very similar to that of $\mathcal{T}$ in Fig. 4, $\mathcal{I}$ does not contain bumps at $\eta_0 \simeq 0.7\pi$. Thus, the bumps in $\mathcal{T}$ are explained solely by $\mathcal{S}$. This suggests that when such

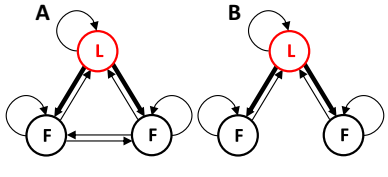

Fig. 2. Three-agent interaction diagrams. (**A**) In model A, a leader influences both followers, both followers influence L, and followers influence each other. (**B**) Model B is similar to model A, but followers cannot influence each other. Weights are asymmetric: w_{LF} (leader to followers) is greater than w_{FL} and w_{FF}. We set $w_{LF} = 4$ and $w_{FL} = w_{FF} = w_{LL} = 1$.

Downloaded from https://www.science.org on March 02, 2022

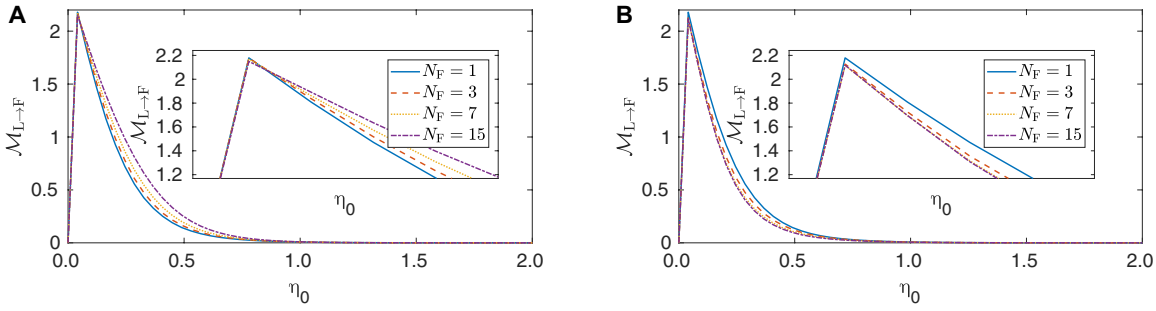

Fig. 3. $\mathcal{M}$ as a function of noise level η_0 (in units of π radians) for models A and B with one leader. Here, N_F = 1 (blue), 3 (red), 7 (yellow), and 15 (purple), where the number of leaders is always one. (**A**) $\mathcal{M}_{L\to F}$ for model A. (**B**) $\mathcal{M}_{L\to F}$ for model B.

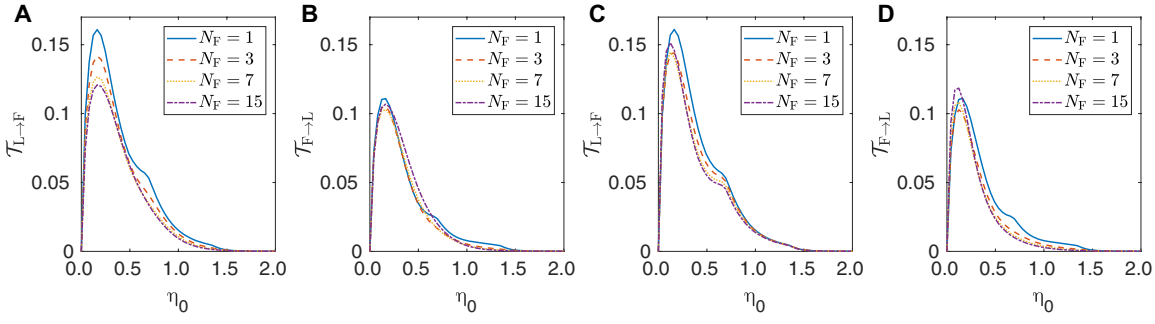

Fig. 4. $\mathcal{T}$ as a function of noise level η_0 (in units of π radians) for models A and B with one leader. Here, N_F = 1 (blue), 3 (red), 7 (yellow), and 15 (purple), where the number of leaders is always one. (**A**) $\mathcal{T}_{L\to F}$ for model A. (**B**) $\mathcal{T}_{F\to L}$ for model A. (**C**) $\mathcal{T}_{L\to F}$ for model B. (**D**) $\mathcal{T}_{F\to L}$ for model B.

a bump exists in the values of $\mathcal{T}$ as a function of noise, this may be attributed to a difference in the location of the peak of $\mathcal{I}$ and that of $\mathcal{S}$. Furthermore, when looking at how this structure changes as a function of the number of following agents, one can deduce whether follower agents mutually interact (i.e., model A) or not (i.e., model B) solely by observing the pairwise trajectories between the leader and one follower. Similar results are also obtained by more simplified binary models (see section S2 and fig. S14). Figure 6 (A to C) shows $\mathcal{S}_{L\to F}$ as a function of η_0 for model A, $\mathcal{S}_{F\to L}$ for model A, and $\mathcal{S}_{L\to F}$ for model B, respectively ($\mathcal{S}_{F\to L}$ for model B is indistinguishable by eye from $\mathcal{S}_{F\to L}$ for model A and therefore not shown here) (see fig. S10). As an overall trend, $\mathcal{S}$ is negligibly small when $\eta_0 < 0.4\pi$, which means that since the configuration of F (L) is not varying by a large amount from time to time, the simultaneous knowledge of L and F does not decrease the uncertainty in F (L) more than knowing the current configuration of L or F alone. At intermediate values of η_0, simultaneous knowledge of L and F becomes relatively more important, and at high values of η_0, the simultaneous knowledge of L and F has no predictive power as the dynamics are dominated by thermal noise. As N_F increases in model A, $\mathcal{S}_{L\to F}$ and $\mathcal{S}_{F\to L}$ both decrease, as the future configuration of F (L) depends on more other agents and relies less on the simultaneous knowledge of L or F alone. Therefore, increasing N_F decreases the likelihood that simultaneously knowing the configuration of F and L has any additional predictive power on L or F. In model B, however, F is not affected by other followers, and therefore, $\mathcal{S}_{L\to F}$ remains largely unchanged as a function of N_F.

Now let us consider the case of multiple leaders in which the motility of follower agents are subject to more than one leader. Figure 7G exemplifies the case of four agents including one leader, while the three followers can interact with one another (model A). Figure 7H exemplifies the same interaction type, but with two leaders and two followers, where the leaders cannot interact with one another but the followers can. Graph representations of cases where the leaders and followers can all mutually interact, leaders can interact with one another but followers cannot, and neither leaders nor followers can interact with one another are shown in fig. S11. The values of $\mathcal{S}$ in these cases are shown in fig. S12, and the results are discussed in section SIF. Here, we study the effect of increasing the number of leaders in model A, where leaders cannot interact with one another but followers can (see Fig. 7, G and H). Figure 7 (A and B) shows $\mathcal{S}_{L\to F}$ and $\mathcal{S}_{F\to L}$, respectively, for the case of four agents, including one leader and three followers (blue) and two leaders and two followers (red). As one may expect, $\mathcal{S}_{L\to F}$ decreases as the number of leaders increases, since the dynamics of each follower results from the two leader agents, reducing synergistic effect between a leader-follower pair in the prediction of the follower motility. In the case of one leader and three followers, a follower is also subject to the interaction of an additional follower instead of the leader; however, since the weight of the follower is less, this does not reduce the synergistic effect as much as the case of two leaders and two followers. Counterintuitively, $\mathcal{S}_{F\to L}$ increases as the number of leaders increases, as shown in Fig. 7 (B and D). Note that keeping the total number of agents fixed, there are fewer followers interacting with a given leader as we increase the number of leaders. It suggests that synergistic effects $\mathcal{S}_{X\to Y}$ decrease as the weighted indegree of the agent Y increases, or, in other words, as more agents "participate" in determining the future of the target agent Y. In Fig. 7 (C and D), where there is a total of eight agents, the same trends are, respectively, seen as the number of leaders is increased; however, the overall

Downloaded from https://www.science.org on March 02, 2022

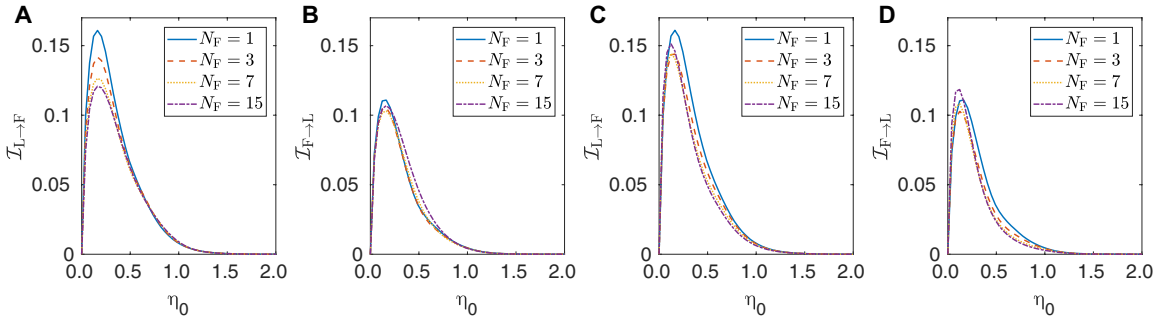

Fig. 5. $\mathcal{I}$ as a function of noise level η_0 (in units of π radians) for models A and B with one leader. Here, N_F = 1 (blue), 3 (red), 7 (yellow), and 15 (purple), where the number of leaders is always one. (**A**) $\mathcal{I}_{L\to F}$ for model A. (**B**) $\mathcal{I}_{F\to L}$ for model A. (**C**) $\mathcal{I}_{L\to F}$ for model B. (**D**) $\mathcal{I}_{F\to L}$ for model B.

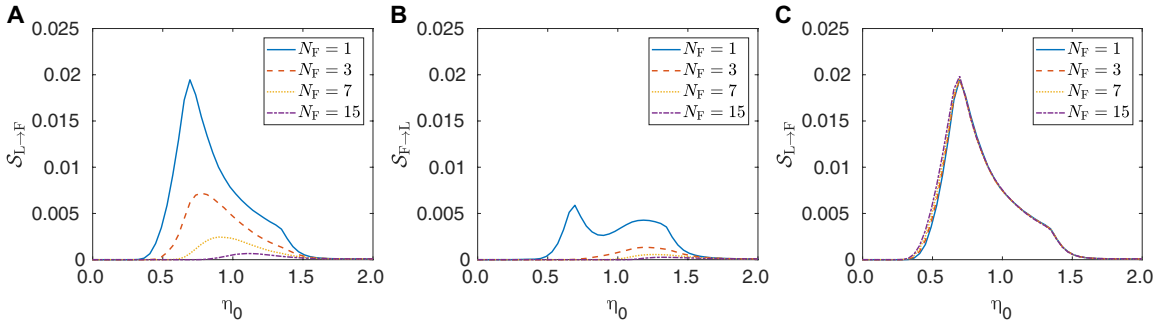

Fig. 6. $\mathcal{S}$ as a function of noise level η_0 (in units of π radians) for models A and B with one leader. Here, N_F = 1 (blue), 3 (red), 7 (yellow), and 15 (purple), where the number of leaders is always one. (**A**) $\mathcal{S}_{L\to F}$ for model A. (**B**) $\mathcal{S}_{F\to L}$ for model A. (**C**) $\mathcal{S}_{L\to F}$ for model B.

values of $\mathcal{S}_{L\to F}$ and $\mathcal{S}_{F\to L}$ are lower than in Fig. 7 (A and B) due to the higher number of agents reducing the synergistic effects. In Fig. 7 (E and F), we keep the number of followers fixed to three and increase the number of leaders. Here, there is no change in $\mathcal{S}_{F\to L}$ as the number of leaders increases, although the total number of agents increases, because the increase in agents does not increase the indegree of L.

DISCUSSION

We investigated a series of model systems based on the Vicsek Model of collective motion to explore the effect of interaction protocols on the distinct modes of information flow. In theory, one would condition on all variables, as well as history, to fully interpret mutual relationships among agents in a collective. At present, this is not practical. Instead, our task was to acquire detailed and correct interpretations under the constraint of limited measurements—specifically, pairwise interactions among agents.

We observed that the intrinsic information between X and Y dominates whenever there is only a link from X to Y and no direct link between Y to X or from Y to itself. However, a small amount of intrinsic information can still be observed when there is no direct link from X to Y, as in the case where X is a follower with memory and Y is a leader. We noted that this was due to the effect of memory. We also found that shared information from X to Y dominates when X and Y are both influenced by a shared history. Synergistic information dominates when present knowledge of X or Y alone cannot predict the future state of Y by itself, but knowing both the present of X and Y simultaneously does. One of the most notable consequences in our analysis of this multiagent system was that decomposing TE into intrinsic and synergistic information flows enabled us to infer whether followers interact with one another in the collective and, more generally, distinguish between cases where an agent interacts mutually with only one or a few other agents, from the cases where many interacting agents are influential. Notably, from that, one can also correctly interpret the "bump" observed in TE as a function of noise level.

Although the concepts of intrinsic, shared, and synergistic information flows apply generally to any system of interacting variables, some of the main limitations of information theory still apply to the measures used in this study. Computing information flow typically requires a discretization of the dataset, which performs a coarse graining of the data (still, these coarse-grained methods have proven more effective than methods based on continuous quantities, such as time-delayed correlation (*12*)). Furthermore, although intrinsic information is an improvement upon TE to rule out the effects of history, it cannot erase all memory effects without conditioning on longer and longer pasts, which, in practice, is not feasible. Notwithstanding these limitations, the decomposition of TE and TDMI into intrinsic, shared, and synergistic modes of information flow provides a marked improvement on the nature of interaction without requiring any additional data.

On the basis of the model systems and their corresponding information flows, one can deduce which information measure is

Downloaded from https://www.science.org on March 02, 2022

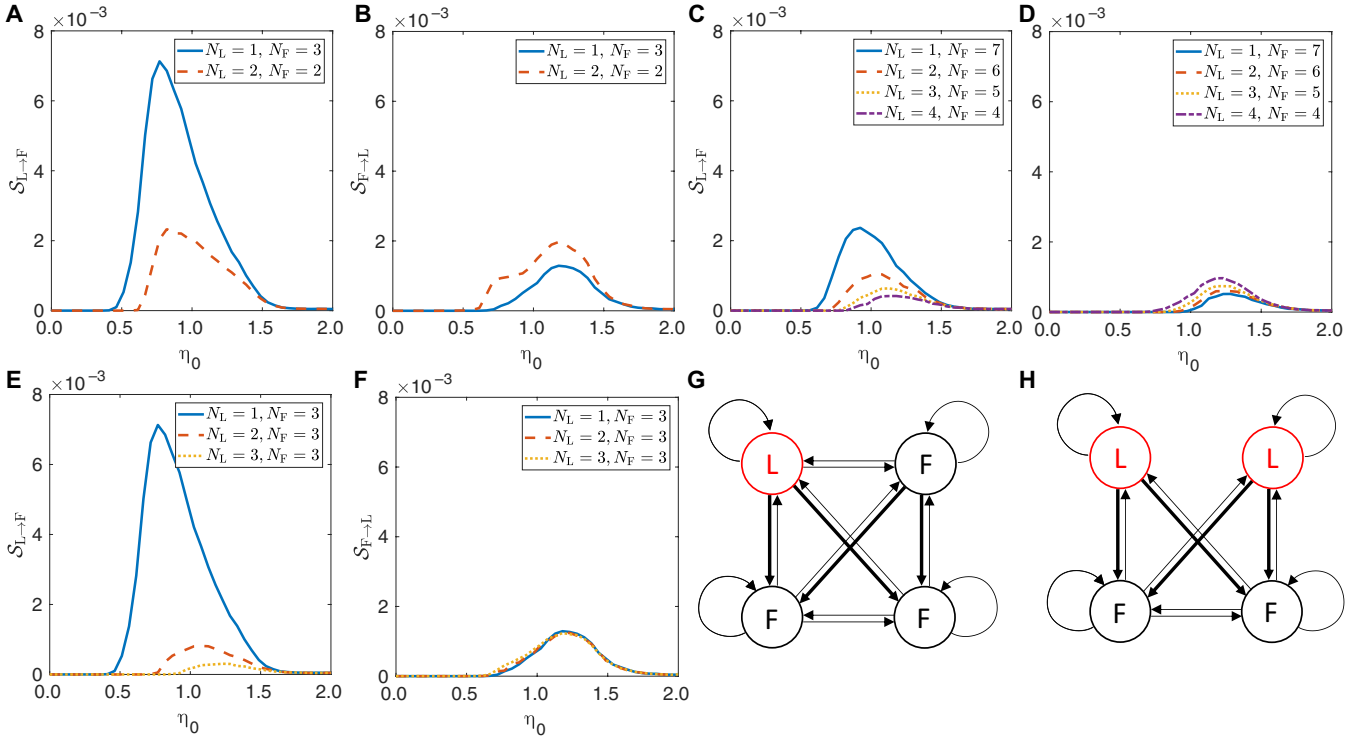

Fig. 7. $\mathcal{S}$ as a function of noise level η_0 (in units of π radians) for model A with different numbers of leaders and followers. (**A** and **B**) $\mathcal{S}_{L\to F}$ (A) and $\mathcal{S}_{L\to F}$ as a function of η_0 (in units of π radians) (B) for four agents with one leader and three followers (blue) and two leaders and two followers (red). (**C** and **D**) $\mathcal{S}_{L\to F}$ (C) and $\mathcal{S}_{F\to L}$. (D) for eight agents with one leader and seven followers (blue), two leaders and six followers (red), three leaders and five followers (yellow), and four leaders and four followers (purple). (**E** and **F**) $\mathcal{S}_{L\to F}$ (E) and $\mathcal{S}_{F\to L}$ (F) with three followers and one leader (blue), two leaders (red), and three leaders (yellow). (**G**) Graph representation of model A, where there is one leader and three followers. (**H**) Graph representation of model A, where there are two leaders and two followers.

more appropriate based on the physical problem being addressed. In leader-follower classification, for example, TE is often used. However, when one does not expect to find substantial synergistic or shared flows, it equals TDMI. The latter is then a better choice since it does not require additional conditioning that increases the dimension of the probability distribution that must be well-sampled. In cases where synergistic flow is dominant, one may consider separating intrinsic and synergistic flows instead of computing just TE. This results in a much richer feature space for classification. In general, computing intrinsic, shared, and synergistic flows should perform better or at least as well as TE and TDMI in classification. Future work will verify these claims and elucidate exactly in which scenarios we expect each mode of information flow to be effective in classifying leaders and followers.

MATERIALS AND METHODS

Defining information flow

In this section, we will construct our measure of intrinsic information flow. We will start with a broader understanding of information flow and then narrow it until we arrive at our goal. To begin, information flow from a time series X to a time series Y must exist in both the behavior of X at time t and the later behavior of Y at time $t + \tau$

$$\mathcal{M}_{X\to Y}(\tau) = I(X_t; Y_{t+\tau}) = \sum_{x_t, y_{t+\tau}} p(x_t, y_{t+\tau}) \log_2 \frac{p(x_t, y_{t+\tau})}{p(x_t)\, p(y_{t+\tau})} \quad (13)$$

a quantity known as the TDMI.

As pointed out by Schreiber (*32*), there are many reasons why X_t and $Y_{t+\tau}$ might share information. First, both X and Y may be synchronized, and so X_t predicts $Y_{t+\tau}$ in the same fashion that Y_t would, and so it would be disingenuous to attribute that shared information to information flow. Similarly, X and Y may be jointly influenced by a third system Z, and so there is no direct, or even indirect, information flow from X to Y. Schreiber referred to these two situations as the two time series being correlated via common history and common input signals and proposed discounting these influences from the TDMI via conditioning

$$\mathcal{T}_{X\to Y}(\tau) = I(X_t; Y_{t+\tau} \mid Y_t) = \sum_{x_t, y_t, y_{t+\tau}} p(y_{t+\tau}, y_t, x_t) \log_2 \frac{p(y_{t+\tau} \mid y_t, x_t)}{p(y_{t+\tau} \mid y_t)} \quad (14)$$

a quantity known as the TE. This does overcome the stated weakness of the TDMI by conditioning on Y_t, thus not including the information shared by X_t and $Y_{t+\tau}$, which also exists in Y_t. The TE can also be modified to discount the information also in a simultaneous third variable

$$I(X_t; Y_{t+\tau} \mid Y_t, Z_t) = \sum_{x_t, y_t, z_t, y_{t+\tau}} p(y_{t+\tau}, y_t, x_t, z_t) \log_2 \frac{p(y_{t+\tau} \mid y_t, x_t, z_t)}{p(y_{t+\tau} \mid y_t, z_t)} \quad (15)$$

Conditioning on variables, however, is not a purely subtractive operation. That is, the following relation does not necessarily hold

$$I(X; Y \mid Z) \le I(X; Y) \quad (16)$$

Downloaded from https://www.science.org on March 02, 2022

Rather, conditioning can increase the information shared by two variables. These phenomena are known as conditional dependence (*51*) and are perhaps best exemplified by the following distribution, where X, Y, and Z are binary random variables, and the events in which an even number of them take on the value 1 have equal probability of $^1/_4$, and the probability of $X = Y = Z = 0$ is also $^1/_4$:

Xor			
X	***Y***	***Z***	**Pr**
0	0	0	$\frac{1}{4}$
0	1	1	$\frac{1}{4}$
1	0	1	$\frac{1}{4}$
1	1	0	$\frac{1}{4}$

In this distribution, any pair of variables is independent

$$I(X;Y) = I(X;Z) = I(Y;Z) = 0 \tag{17}$$

yet each of those pairs conditioned on the third variable is highly correlated

$$I(X;Y\,|\,Z) = I(X;Z\,|\,Y) = I(Y;Z\,|\,X) = 1 \tag{18}$$

This is because knowing the value of, for example, X, does not allow us to infer the value of Y, but if conditioned on (again, for example) $Z = 0$, suddenly, we know that whatever value X takes, Y must as well. This is conditional dependence in its purest form: X and Y are independent, but given Z, they are perfectly correlated.

This brings us back to the TE. As it is based on a particular conditional mutual information, conditioning on Y_t can induce an apparent correlation between X_t and $Y_{t+\tau}$, which does not exist without it. To overcome this weakness in the TE, it was proposed to take a step back and consider the problem from a slightly different perspective (*33*). We seek an operational understanding of the information shared by X_t and $Y_{t+\tau}$, removing the influences of common history and common input signals without introducing other forms of correlation. To do this, we appeal to the cryptographic flow ansatz (*33*), which states that intrinsic information flow exists when X_t and $Y_{t+\tau}$ can agree upon a secret, while Y_t acts as an eavesdropper, and furthermore, that the intrinsic information flow is quantified as the rate of secret sharing existing between the two (see section S3). In essence, this means that information shared by X_t and $Y_{t+\tau}$ can only be definitively attributed to flow from X_t to $Y_{t+\tau}$ if there is no way that information can be reconstructed or derived by Y_t.

To practically apply the cryptographic flow ansatz, we use a relatively easily computable upper bound, termed as IMI

$$\mathcal{I}_{X\to Y}(\tau) = \inf_{p(\overline{Y}_t=\bar{y}_t\,|\,Y_t=y_t)} I(X_t;Y_{t+\tau}\,|\,\overline{Y}_t) \tag{19}$$

Effectively, this bound simply states that the information shared by X_t and $Y_{t+\tau}$, which is inaccessible to Y_t, is bound from above by the conditional mutual information between X_t and $Y_{t+\tau}$ given all possible variables that can be constructed from $\overline{Y}_t$.

The calculation of the IMI, while not trivial, is not particularly difficult. While the optimization over $p(\overline{Y}_t\,|\,Y_t)$ is not convex, the optimization space is finite because the cardinality of $\overline{Y}_t$ can be bound by the cardinality of Y_t, $|Y_t|$ (*52*). The object of optimization is then a $|Y_t|\times|Y_t|$ row-stochastic matrix, where the i,jth entry is $p(\overline{Y}_t = \bar{y}_j\,|\,Y_t = y_i)$. Global optimization techniques, such as basin hopping, can then be used to find the global minimum. In basin hopping, an initial condition is proposed, and the local minima are found through standard gradient-based techniques; then a step in the optimization space is taken, and the local minima are found again. This is repeated some number of times, and the least of the found local minima is presumed to be the global minimum. This is the technique used in the dit information theory package (*53*), which was used to perform the calculations in this manuscript.

Briefly, this measure builds upon the TE, producing a previously unidentified metric that comes substantially closer to the TE's stated goal of removing the effects of common history and common input signals but without introducing the possibility of conditional dependence. This is accomplished by appealing to the field of information-theoretic cryptography and drawing parallels between secret key agreement and the scientific issue of attributing information present in $Y_{t+\tau}$ to X_t and X_t alone.

IMI is a sharp bound on the secret key agreement rate. Note that only in particular cases have rates been derived. However, in these cases, IMI and secret key agreement rate coincide (*48*, *53*). That said, probability distributions with IMI greater than the secret key agreement rate can be constructed. Again, only a few constructions are known (*54*, *55*). In these situations, tighter bounds on secret key agreement rate exist. However, they are more challenging to calculate (*54*, *55*).

Computing information flow measures

The computation of $\mathcal{T}$, $\mathcal{M}$, $\mathcal{S}$, and σ were performed as follows. First, the orientations θ_L and θ_F are computed as described by Eq. 12 up to time $T = 2\times10^6$ for 20 sets of initial conditions. The values of θ_L and θ_F are then discretized by binning them into six bins on the interval $[0,2\pi]$ [the use of six symbols was found to be sufficient to differentiate the behaviors of L and F while maintaining a computationally feasible number of sequences to be sampled to compute information measures; see (*12*, *24*)]. For each set of initial conditions, the joint probability distribution of $p(x(t), y(t), y(t+1))$, from the time series, where $x(t)$ and $y(t)$ are the discretized forms of θ_i, where x and y can be either L or F. $p(x(t), y(t), y(t+1))$ is computed by counting the occurrences of each of the 6^3 possible combinations of $(x(t), y(t), y(t+1))$ and dividing by the total length of the time series minus 1, $2\times10^6 - 1$. Once the probability distributions are computed, $\mathcal{T}$ is computed by plugging them into the equation $\mathcal{T} = \sum_{y_{t+\tau},y_t,x_t} p(y_{t+\tau},y_t,x_t)\log_2\frac{p(y_{t+\tau}|y_t,x_t)}{p(y_{t+\tau}|y_t)}$ (*32*), where $\tau = 1$. Likewise, $\mathcal{M}$ is computed using the formula $\mathcal{M} = \sum_{x_t,y_{t+\tau}} p(x_t,y_{t+\tau})\log_2\frac{p(x_t,y_{t+\tau})}{p(x_t)\,p(y_{t+\tau})}$. The calculation of $\mathcal{I}$ is described in the "Defining information flow" section. Last, σ and $\mathcal{S}$ are computed directly using Eqs. 7 and 8 ($\sigma = \mathcal{M} - \mathcal{I}$ and $\mathcal{S} = \mathcal{T} - \mathcal{I}$).

In practice, one does not need to use longer time series to compute the modes of information flow than those required for the computations of TE or TDMI. To further confirm the statistical significance of our results on the decomposed mode of information flow, we have performed a surrogate test (*56*, *57*). This uses surrogate time series by swapping a pair of two time series of two interacting agents in the set of realizations, which preserves all statistical properties of the nonlinear dynamics of the individual agents but spoils causal relationship if it exists. We confirmed that our conclusions are statistically significant (see the "Surrogate test" section and Fig. 8).

Downloaded from https://www.science.org on March 02, 2022

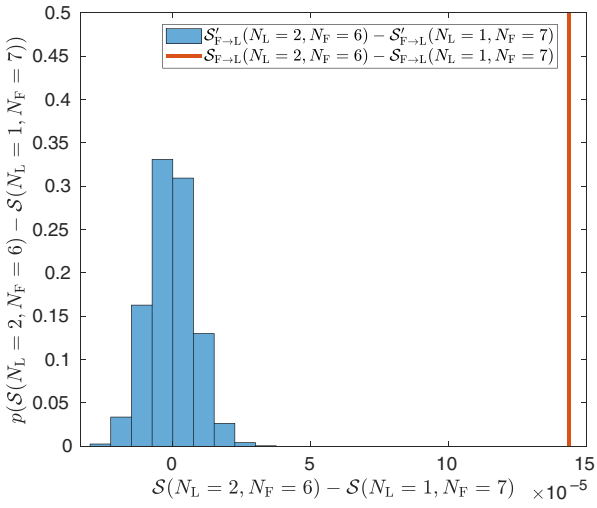

Fig. 8. Surrogate and actual distributions used for surrogate testing. (Blue) Distribution of surrogate values $\mathcal{S}'_{F\to L}(N_L = 2, N_F = 6) - \mathcal{S}_{F\to L}(N_L = 1, N_F = 7)$. (Red) the actual value of $\mathcal{S}_{F\to L}(N_L = 2, N_F = 6) - \mathcal{S}_{F\to L}(N_L = 1, N_F = 7)$ averaged over 20 realizations for $\eta_0 = 1.1\pi$. The values of $\mathcal{S}_{F\to L}(N_L = 2, N_F = 6)$ and $\mathcal{S}_{F\to L}(N_L = 1, N_F = 7)$ for η_0 ranging from 0 to 2π are shown in Fig. 7D.

Details of the modified Vicsek models

Here, we describe the details of the modified Vicsek model. To iterate their trajectories, the following formula is used (*12*)

$$\langle\theta(t+1)\rangle_{R,\mathbf{w},\vec{\mathbf{r}}_i^t} = \arctan\left[\sum_j' w_{ij}\sin\theta_j(t)/\sum_j' w_{ij}\cos\theta_j(t)\right] \quad (20)$$

Here $\sum_j'$ sums over all j satisfying $|\vec{\mathbf{r}}_i^t - \vec{\mathbf{r}}_j^t| \le R$. In this, $\mathbf{w}$ is a nonnegative asymmetric matrix whose w_{ij} element determines the interaction strength that particle i exerts on particle j. $w_{ij} > w_{ji}$ whenever particle i is a leader and particle j is a follower in our setting. Positions at the initial time $t = 1$ are chosen randomly from a uniform distribution within the box of length $L = 10$, and orientations are chosen randomly from a uniform distribution on the interval $[0,2\pi)$. The interaction radius R is set to $R = 3$. Positions are updated using Eq. 11, and the orientations θ_F and θ_L are updated using Eq. 12. All simulations are coded in MATLAB.

Surrogate test

To validate the statistical significance of the values of $\mathcal{M}$, $\mathcal{T}$, $\mathcal{I}$, σ, and $\mathcal{S}$ reported here and in the Supplementary Materials, we perform a permutation test using surrogate data following (*56*). When performing the statistical test on $\mathcal{S}_{F\to L}$, for example, we first propose the null hypothesis that there is no significant amount of $\mathcal{S}_{F\to L}$. Then, we create a distribution using surrogate data that preserves the dynamical and statistical properties of $F(t)$ and $L(t)$ while spoiling causality between them if it exists. In random shuffle surrogate analysis, the time series $F(t)$ is randomly shuffled, meaning $F(t)$ at each time t is replaced by a randomly chosen $F(t')$ from another time t'; however, this method does not preserve dynamical properties of $F(t)$, such as the dependency of $F(t + 1)$ on $F(t)$. Fourier transform–based surrogates, such as iterative amplitude adjusted Fourier transform (*57*), are often used to preserve both the linear correlation in time (i.e., power spectrum) and the frequency distribution of the original time series. Here, desired surrogate data to unveil statistical significance for observed information modes in TE and TDMI are those preserving nonlinear dynamics and frequency distribution intrinsic to each stochastic variable of leaders and followers, θ_{L_i} and θ_{F_j}, but spoiling possible causal relation in the computation of information mode decomposition.

We first performed a set of simulations for n realizations, each of which the initial positions of agents are taken randomly from a uniform distribution within a 10 by 10 box with a periodic boundary condition (the interaction radius R is equal to 3), and their orientations from a uniform distribution on the interval $[0,2\pi)$, respectively. The kth realization of the ith leader's and the jth follower's dynamics are denoted as $F_i^{(k)}(t)$ and $L_j^{(k)}(t)$ $(1 \le i \le N_L, 1 \le j \le N_F)$, respectively, where N_L and N_F are the number of leaders and that of followers in the system, respectively.

For example, to carry out the surrogate analysis on synergistic information $\mathcal{S}_{F_j\to L_i} = \frac{1}{n}\sum_{k=1}^{n}\mathcal{S}_{F_j^{(k)}\to L_i^{(k)}}$, we simply swapped the (same length) time series of the kth realization $F_j^{(k)}(t)$ or $L_i^{(k)}(t)$ to the other k'th realization $F_{j'}^{(k')}(t)$ or $L_{i'}^{(k')}(t)$, which preserves the nonlinear dynamical and statistical properties of leaders and followers, while any causal relation is erased simply because of different realization. Namely, the surrogate "synergistic information" is defined by $\mathcal{S}'_{F_j\to L_i} \equiv \left\{\mathcal{S}_{F_j^{(k)}\to L_{i'}^{(k')}}, \mathcal{S}_{F_{j'}^{(k')}\to L_i^{(k)}}\right\}$. Here, i, i', j, and j' were chosen arbitrarily from the set of N_L leaders, $\{L_i\}$, and that of N_F followers, $\{F_j\}$, in kth and k'th realizations (e.g., i and i' over different realizations are not necessarily identical among the same kind of agents and randomly taken from $\{L_i\}$ over different realizations). It was confirmed that decomposed information modes including $\mathcal{S}_{F_j\to L_i}$ were statistically indistinguishable among different choices of i and j across (n = 20) realizations here, which implies that the finite length of trajectories ($T = 2 \times 10^6$) and the number of 20 realizations were enough to characterize the general properties of decomposed information modes between leader(s) and follower(s) in the models. Thus, here, we simply omitted the notation of subscripts i and j from L_i and F_j like $\mathcal{S}_{F\to L}$ and $\mathcal{S}'_{F\to L}$. If the null hypothesis, i.e., no synergistic information exists, is true, it means that the value of $\mathcal{S}_{F\to L}$ should not be indistinguishable statistically from $\mathcal{S}'_{F\to L}$. One can quantify this by setting a false-positive error rate α by using the distribution of the null hypothesis $p(\mathcal{S}'_{F\to L})$ so that the value of $\mathcal{S}_{F\to L}$ is fallen into the domain of $1 - \alpha$. If it is false, then the value of $\mathcal{S}_{F\to L}$ would appear as an outlier outside of the domain.

Here, for example, Fig. 8 demonstrates the surrogate test with n = 10,000 (= 100 × 100) surrogate trials to verify the statistical significance of the difference between $\mathcal{S}_{F\to L}$, where $N_L = 2$ and $N_F = 6$ and $\mathcal{S}_{F\to L}$, where $N_L = 1$ and $N_F = 7$ for the multiagent Vicsek model A [see $\mathcal{S}_{F\to L}(N_L = 2, N_F = 6) - \mathcal{S}_{F\to L}(N_L = 1, N_F = 7)$ in Fig. 7D]. The figure manifests that $\mathcal{S}_{F\to L}(N_L = 2, N_F = 6) - \mathcal{S}_{F\to L}(N_L = 1, N_F = 7)$ are statistically significant with significance level $\alpha < 0.01$ (i.e., no false positives over 10,000 surrogate trials), implying that our interpretation of "the more the N_L, the larger $\mathcal{S}_{F\to L}$ under $N_L + N_F$ being invariant" in Fig. 7D was validated.

Likewise, we confirmed statistical significances of information mode decomposition here, especially when the amounts of the quantities are marginally small.

It may be noted for a limited number of trajectories, such as in experimental data, that the nearest-neighbor permutation test (*58*) can be an alternative method of surrogate analysis, in which the time series $F(t)$ at each time t is replaced by a randomly chosen $F(t')$ at any time t' $(\neq t)$ where $|L(t) - L(t')|$ is the smallest (nearest). This preserves the marginal distributions of $F(t)$ and $L(t)$, and the joint probability distribution $p(F(t), L(t))$ between $F(t)$ and $L(t)$ if the

Downloaded from https://www.science.org on March 02, 2022

nearest neighbors chosen are close enough to guarantee $L(t) \approx L(t')$ while spoiling $p(F(t), L(t), L(t+1))$ between $F(t)$ and $L(t+1)$.

SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at https://science.org/doi/10.1126/sciadv.abj1720

View/request a protocol for this paper from *Bio-protocol*.

REFERENCES AND NOTES

1. K. Copenhagen, D. A. Quint, A. Gopinathan, Self-organized sorting limits behavioral variability in swarms. *Sci. Rep.* **6**, 31808 (2016).
2. H. Hornischer, S. Herminghaus, M. G. Mazza, Structural transition in the collective behavior of cognitive agents. *Sci. Rep.* **9**, 12477 (2019).
3. M. Nagy, Z. Ákos, D. Biro, T. Vicsek, Hierarchical group dynamics in pigeon flocks. *Nature* **464**, 890–893 (2010).
4. M. Porfiri, Inferring causal relationships in zebrafish-robot interactions through transfer entropy: A small lure to catch a big fish. *Anim. Cogn.* **5**, 341–367 (2018).
5. S. Butail, V. Mwaffo, M. Porfiri, Model-free information-theoretic approach to infer leadership in pairs of zebrafish. *Phys. Rev. E* **93**, 042411 (2016).
6. C. J. Torney, M. Lamont, L. Debell, R. J. Angohiatok, L.-M. Leclerc, A. M. Berdahl, Inferring the rules of social interaction in migrating caribou. *Philos. Trans. R. Soc. B* **373**, 20170385 (2018).
7. A. Strandburg-Peshkin, D. R. Farine, I. D. Couzin, M. C. Crofoot, Shared decision-making drives collective movement in wild baboons. *Science* **348**, 1358–1361 (2015).
8. M. Reffay, L. Petitjean, S. Coscoy, E. Grasland-Mongrain, F. Amblard, A. Buguin, P. Silberzan, Orientation and polarity in collectively migrating cell structures: Statics and dynamics. *Biophys. J.* **100**, 2566–2575 (2011).
9. M. Reffay, M.-C. Parrini, O. Cochet-Escartin, B. Ladoux, A. Buguin, S. Coscoy, F. Amblard, J. Camonis, P. Silberzan, Interplay of rhoa and mechanical forces in collective cell migration driven by leader cells. *Nat. Cell Biol.* **16**, 217–223 (2014).
10. N. Yamaguchi, T. Mizutani, K. Kawabata, H. Haga, Leader cells regulate collective cell migration via rac activation in the downstream signaling of integrin β1 and pi3k. *Sci. Rep.* **5**, 7656 (2015).
11. J. Garland, A. M. Berdahl, J. Sun, E. M. Bollt, Anatomy of leadership in collective behaviour. *Chaos* **28**, 075308 (2018).
12. U. S. Basak, S. Sattari, K. Horikawa, T. Komatsuzaki, Inferring domain of interactions among particles from ensemble of trajectories. *Phys. Rev. E* **102**, 012404 (2020).
13. V. Mwaffo, S. Butail, M. Porfiri, Analysis of pairwise interactions in a maximum likelihood sense to identify leaders in a group. *Front. Robot. AI* **4**, 35 (2017).
14. V. Mwaffo, J. Keshavan, T. L. Hedrick, S. Humbert, Detecting intermittent switching leadership in coupled dynamical systems. *Sci. Rep.* **8**, 10338 (2018).
15. V. Mwaffo, J. Keshavan, T. Hedrick, S. Humbert, A data-driven method to dissect the dynamics of the causal influence in complex dynamical systems, in *COMPENG 2018* (IEEE, 2018), pp. 1–5.
16. K. Takamizawa, M. Kawasaki, Transfer entropy for synchronized behavior estimation of interpersonal relationships in human communication: Identifying leaders or followers. *Sci. Rep.* **9**, 10960 (2019).
17. T. Liu, L. Ungar, K. Kording, Quantifying causality in data science with quasi-experiments. *Nat. Comp. Sci.* **1**, 24–32 (2021).
18. S. L. Bressler, A. K. Seth, Wiener–granger causality: A well established methodology. *Neuroimage* **58**, 323–329 (2011).
19. J. Borge-Holthoefer, N. Perra, B. Gonçalves, S. González-Bailón, A. Arenas, Y. Moreno, A. Vespignani, The dynamics of information-driven coordination phenomena: A transfer entropy analysis. *Sci. Adv.* **2**, e1501158 (2016).
20. F. A. Lavergne, H. Wendehenne, T. Bäuerle, C. Bechinger, Group formation and cohesion of active particles with visual perception–dependent motility. *Science* **364**, 70–74 (2019).
21. J. L. Servadio, M. Convertino, Optimal information networks: Application for data-driven integrated health in populations. *Sci. Adv.* **4**, e1701088 (2018).
22. M. Nitzan, J. Casadiego, M. Timme, Revealing physical interaction networks from statistics of collective dynamics. *Sci. Adv.* **3**, e1600396 (2017).
23. J. F. Ramirez-Villegas, M. Besserve, Y. Murayama, H. C. Evrard, A. Oeltermann, N. K. Logothetis, Coupling of hippocampal theta and ripples with pontogeniculooccipital waves. *Nature* **589**, 96–102 (2021).
24. U. S. Basak, S. Sattari, M. M. Hossain, K. Horikawa, T. Komatsuzaki, An information-theoretic approach to infer the underlying interaction domain among elements from finite length trajectories in a noisy environment. *J. Chem. Phys.* **154**, 034901 (2021).
25. N. Orange, N. Abaid, A transfer entropy analysis of leader-follower interactions in flying bats. *Eur. Phys. J. Spec. Top.* **224**, 3279–3293 (2015).
26. W. M. Lord, J. Sun, N. T. Ouellette, E. M. Bollt, Inference of causal information flow in collective animal behavior. *IEEE Trans. Mol. Biol. Multi-Scale Commun.* **2**, 107–116 (2016).
27. J. Jeong, J. C. Gore, B. S. Peterson, Mutual information analysis of the EEG in patients with Alzheimer's disease. *Clin. Neurophysiol.* **112**, 827–835 (2001).
28. J. Sun, D. Taylor, E. M. Bollt, Causal network inference by optimal causation entropy. *SIAM J. Appl. Dyn. Syst.* **14**, 73–106 (2015).
29. J. Brown, T. Bossomaier, L. Barnett, Information flow in finite flocks. *Sci. Rep.* **10**, 3837 (2020).
30. L. Barnett, J. T. Lizier, M. Harré, A. K. Seth, T. Bossomaier, Information flow in a kinetic Ising model peaks in the disordered phase. *Phys. Rev. Lett.* **111**, 177203 (2013).
31. S. Roy, Quantifying interactions among car drivers using information theory. *Chaos* **30**, 113125 (2020).
32. T. Schreiber, Measuring information transfer. *Phys. Rev. Lett.* **85**, 461–464 (2000).
33. R. G. James, B. D. M. Ayala, B. Zakirov, J. P. Crutchfield, Modes of information flow. arXiv:1808.06273 [cond-mat.stat-mech] (2018).
34. T. M. Cover, J. A. Thomas, *Elements of Information Theory* (Wiley-Interscience, ed. 2, 2006).
35. T. Vicsek, A. Czirók, E. Ben-Jacob, I. Cohen, O. Shochet, Novel type of phase transition in a system of self-driven particles. *Phys. Rev. Lett.* **75**, 1226–1229 (1995).
36. R. Vicente, M. Wibral, M. Lindner, G. Pipa, Transfer entropy–a model-free measure of effective connectivity for the neurosciences. *J. Comput. Neurosci.* **30**, 45–67 (2011).
37. R. E. Spinney, M. Prokopenko, J. T. Lizier, Transfer entropy in continuous time, with applications to jump and neural spiking processes. *Phys. Rev. E* **95**, 032319 (2017).
38. P. Wollstadt, M. Martínez-Zarzuela, R. Vicente, F. J. Díaz-Pernas, M. Wibral, Efficient transfer entropy analysis of non-stationary neural time series. *PLOS ONE* **9**, e102833 (2014).
39. M. Wibral, B. Rahm, M. Rieder, M. Lindner, R. Vicente, J. Kaiser, Transfer entropy in magnetoencephalographic data: Quantifying information flow in cortical and cerebellar networks. *Prog. Biophys. Mol. Biol.* **105**, 80–97 (2011).
40. M. Wibral, R. Vicente, M. Lindner, Transfer entropy in neuroscience, in *Directed Information Measures in Neuroscience* (Springer, 2014), pp. 3–36.
41. M. Bauer, J. W. Cox, M. H. Caveness, J. J. Downs, N. F. Thornhill, Finding the direction of disturbance propagation in a chemical process using transfer entropy. *IEEE Trans. Control Syst. Technol.* **15**, 12–21 (2006).
42. S. Nakayama, M. R. Marín, M. Camacho, M. Porfiri, Plasticity in leader–follower roles in human teams. *Sci. Rep.* **7**, 14562 (2017).
43. M. Porfiri, R. R. Sattanapalle, S. Nakayama, J. Macinko, R. Sipahi, Media coverage and firearm acquisition in the aftermath of a mass shooting. *Nat. Hum. Behav.* **3**, 913–921 (2019).
44. A. Delgado-Bonal, A. Marshak, Y. Yang, D. Holdaway, Analyzing changes in the complexity of climate in the last four decades using MERRA-2 radiation data. *Sci. Rep.* **10**, 922 (2020).
45. T. Gerken, B. L. Ruddell, R. Yu, P. C. Stoy, D. T. Drewry, Robust observations of land-to-atmosphere feedbacks using the information flows of FLUXNET. *npj Clim. Atmos. Sci.* **2**, 37 (2019).
46. S. Campuzano, A. De Santis, F. J. Pavón-Carrasco, M. L. Osete, E. Qamili, New perspectives in the study of the Earth's magnetic field and climate connection: The use of transfer entropy. *PLOS ONE* **13**, e0207270 (2018).
47. R. G. James, N. Barnett, J. P. Crutchfield, Information flows? A critique of transfer entropies. *Phys. Rev. Lett.* **116**, 238701 (2016).
48. U. M. Maurer, S. Wolf, Unconditionally secure key agreement and the intrinsic conditional information. *IEEE Trans. Inf. Theory* **45**, 499–514 (1999).
49. R. G. James, J. Emenheiser, J. P. Crutchfield, Unique information and secret key agreement. *Entropy* **21**, 12 (2019).
50. A. Kaiser, T. Schreiber, Information transfer in continuous processes. *Physica D* **166**, 43–62 (2002).
51. D. Husmeier, *Probabilistic Modeling in Bioinformatics and Medical Informatics* (Springer, 2005).
52. M. Christandl, R. Renner, S. Wolf, A property of the intrinsic mutual information, in *ISIT 2003 Proceed* (IEEE, 2003), pp. 258–258.
53. R. James, C. Ellison, J. Crutchfield, "dit": A Python package for discrete information theory. *J. Open Source Softw.* **3**, 738 (2018).
54. R. Renner, J. Skripsky, S. Wolf, New measure for conditional mutual information and its properties, in *ISIT 2003 Proceed* (IEEE, 2003), pp. 259–259.
55. A. A. Gohari, V. Anantharam, Information-theoretic key agreement of multiple terminals–Part I. *IEEE Trans. Inf. Theory* **56**, 3973–3996 (2010).
56. J. T. Lizier, JIDT: An information-theoretic toolkit for studying the dynamics of complex systems. *Front. Rob. AI* **1**, 11 (2014).
57. G. Lancaster, D. Iatsenko, A. Pidde, V. Ticcinelli, A. Stefanovska, Surrogate data for hypothesis testing of physical systems. *Phys. Rep.* **748**, 1–60 (2018).
58. J. Runge, Causal network reconstruction from time series: From theoretical assumptions to practical estimation. *Chaos* **28**, 075310 (2018).

Downloaded from https://www.science.org on March 02, 2022

59. W. Diffie, New directions in cryptography. *IEEE Trans. Inform. Theory* **22**, 644–654 (1976).
60. S. Y. Yan, *Number Theory for Computing* (Springer Science and Business Media, 2002).
61. D. Adrian, K. Bhargavan, Z. Durumeric, P. Gaudry, M. Green, J. A. Halderman, N. Heninger, D. Springall, E. Thomé, L. Valenta, B. VanderSloot, E. Wustrow, S. Zannela-Béguelin, P. Zimmerman, Imperfect forward secrecy: How Diffie-Hellman fails in practice. *Commun. ACM* **62**, 106–114 (2018).
62. U. M. Maurer, Secret key agreement by public discussion from common information. *IEEE Trans. Inf. Theory* **39**, 733–742 (1993).

Acknowledgments: We thank K. Horikawa, J. Green, M. Toda, H. Teramoto, and K. Tabata for valuable discussions. T.K. and J.P.C. thank the Telluride Science Research Center for its hospitality during visits. T.K. and L.W.P. thank MANABIYA system provided by WPI-ICReDD, Hokkaido University. **Funding:** This work was supported by a Grant-in-Aid for Scientific Research on Innovative Areas "Singularity Biology (No.8007)"(18H05413), MEXT, and by JSPS (no. 25287105 and 25650044 to T.K.), JST/CREST (no. JPMJCR1662 to T.K.), and Cooperative Research Program of the Network Joint Research Center for Materials and Devices (no. 20191062-01 to T.K.). It is also based on work supported by, or in part by, the U.S. Army Research Laboratory and the U.S. Army Research Office under grants W911NF-18-1-0028 and W911NF-21-1-0048 to J.P.C. **Author contributions:** All authors contributed to the conceptual and theoretical aspects of the study and wrote the manuscript. U.S.B. and R.G.J. performed initial exploratory numerical experiments, which the concepts of the manuscript are based on. S.S. carried out the numerical simulations and created the figures in the manuscript. L.W.P. performed the statistical tests. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. The code used to generate all simulations is available at https://doi.org/10.5281/zenodo.5722850. Updates to the simulation code after the publication of this manuscript will be available at https://github.com/sulimon510/Agent-Based-Modeling.

Submitted 25 April 2021
Accepted 20 December 2021
Published 9 February 2022
10.1126/sciadv.abj1720

Downloaded from https://www.science.org on March 02, 2022

Science Advances

Modes of information flow in collective cohesion

Sulimon SattariUdoy S. BasakRyan G. JamesLouis W. PerrinJames P. CrutchfieldTamiki Komatsuzaki

Sci. Adv., 8 (6), eabj1720.

View the article online
https://www.science.org/doi/10.1126/sciadv.abj1720
Permissions
https://www.science.org/help/reprints-and-permissions

Downloaded from https://www.science.org on March 02, 2022

Use of think article is subject to the Terms of service

Science Advances (ISSN) is published by the American Association for the Advancement of Science. 1200 New York Avenue NW, Washington, DC 20005. The title *Science Advances* is a registered trademark of AAAS.
Copyright © 2022 The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original U.S. Government Works. Distributed under a Creative Commons Attribution NonCommercial License 4.0 (CC BY-NC).