

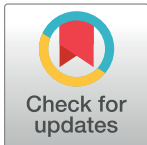
RESEARCH ARTICLE

Finding continuity and discontinuity in fish schools via integrated information theory

Takayuki Niizato¹*, Kotaro Sakamoto²*, Yoh-ichi Mototake³, Hisashi Murakami⁴, Takenori Tomaru⁵, Tomotaro Hoshika¹, Toshiki Fukushima¹

1 Faculty of Engineering, Information and Systems, University of Tsukuba, Tsukuba, Ibaraki, Japan, **2** University of Tsukuba, Leading Graduate School Doctoral Program in Human Biology, Tsukuba, Japan, **3** The Institute of Statistical Mathematics, Tachikawa, Japan, **4** University of Tokyo, Research Center for Advanced Science and Technology, Tokyo, Japan, **5** Department of Computer Science and Engineering, Toyohashi University of Technology, Toyohashi, Japan

* These authors contributed equally to this work.

* t_niizato@yahoo.co.jp (TN); sakamoto@ccs.tsukuba.ac.jp (KS)

Abstract

Collective behaviours are known to be the result of diverse dynamics and are sometimes likened to living systems. Although many studies have revealed the dynamics of various collective behaviours, their main focus has been on the information processing performed by the collective, not on interactions within the collective. For example, the qualitative difference between three and four elements in a system has rarely been investigated. Tononi et al. proposed integrated information theory (IIT) to measure the degree of consciousness Φ . IIT postulates that the amount of information loss caused by the minimum information partition is equivalent to the degree of information integration in the system. This measure is not only useful for estimating the degree of consciousness but can also be applied to more general network systems. Here, we obtained two main results from the application of IIT (in particular, IIT 3.0) to the analysis of real fish schools (*Plecoglossus altivelis*). First, we observed that the discontinuity on $\langle \Phi(M) \rangle$ distributions emerges for a school of four or more fish. This transition was not observed by measuring the mutual information or the sum of the transfer entropy. We also analysed the IIT on Boids simulations with respect to different coupling strengths; however, the results of the Boids model were found to be quite different from those of real fish. Second, we found a correlation between this discontinuity and the emergence of leadership. We discriminate leadership in this paper from its traditional meaning (e.g. defined by transfer entropy) because IIT-induced leadership refers not to group behaviour, as in other methods, but the degree of autonomy (i.e. group integrity). These results suggest that integrated information Φ can reveal the emergence of a new type of leadership which cannot be observed using other measures.

OPEN ACCESS

Citation: Niizato T, Sakamoto K, Mototake Y-i, Murakami H, Tomaru T, Hoshika T, et al. (2020) Finding continuity and discontinuity in fish schools via integrated information theory. PLoS ONE 15(2): e0229573. <https://doi.org/10.1371/journal.pone.0229573>

Editor: Bryan C Daniels, Arizona State University & Santa Fe Institute, UNITED STATES

Received: March 7, 2019

Accepted: February 10, 2020

Published: February 27, 2020

Copyright: © 2020 Niizato et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the manuscript and its Supporting Information files.

Funding: The authors declare that they received no specific funding for this work.

Competing interests: The authors declare that no competing interests exist.

Introduction

Collective behaviours in nature are emergent properties produced through the local interactions of self-organising individuals. Such behaviours include swarming [1–6], fish schooling

[7–11], bird flocking [12–17], or high-level cognitive functions arising from ‘bottom-up’ neural networks [18–21]. These systems of many interacting elements can achieve optimal information processing capabilities when poised at the critical boundary separating chaos from order [22–24]. Several analyses have demonstrated individuals responding swiftly as a collective to changing environments [14–16] and that a group can achieve relatively good decision-making [25–27]. Conflicts among individuals do not necessarily lead to group disruption and can instead provide valuable insight into what collective responses are effective [28, 29]. The unity of collective behaviours remains an unsolved question of nature [30–32], because the interactions are hidden, whereas the resultant actions are observable.

Self-organised criticality (SOC) is a valuable metaphor for interpreting complex collective behaviours. Flexibility and robustness can be achieved collectively when a group is in the intermediate state between order and disorder [33–38]. For instance, the external perturbations of collective systems in SOC models optimise the effective correlation range of each individual to efficiently transfer information [13–16]. However, the SOC models are unreliable for small groups, because individual interactions are less likely to be homogeneous [39, 40]. In fish, for example, groups of two and three appear to exhibit differential between-individual interactions [41, 42]. Therefore, many researchers have examined information transfer (or causal relationships) among individuals in small groups [43–45], often employing local transfer entropy (TE) [46–51]. For example, the transfer of misinformation can happen in schools with five fish when the whole group changes direction [45]. Animals could potentially use active information storage to predict the timing of nontrivial information transfer [52, 53]. However, despite capturing some ‘what the system does’ aspects of collective behaviour systems (that is, ‘happening’: the actual behaviour), SOC models provide little insight on ‘what the system is’ (or ‘being’: the potential dynamical complexity). A detailed discussion on the difference between ‘being’ and ‘happening’ can be found in [54] and [55].

In contrast, integrated information theory (IIT) may be a good metric for ‘what the system is’. Proposed by Tononi and Sporns [56, 57] in 2001, IIT was originally developed to quantify consciousness from brain activity [58, 59]. In principle, the core concept of IIT is to define integrated information as the degree of information loss (or increase in uncertainty [60]) caused by a given partition of the system [59]. Integrated information is designed to quantify the holistic amount of cause–effect power possessed by a system that goes beyond and above the sum of its parts. The latest version of IIT (IIT 3.0 [61]) postulates that an arbitrary subset of elements of the system is a ‘mechanism’ if its intrinsic cause–effect power, i.e. the ability to determine the future and constrain the past states of other arbitrary parts (‘purviews’), is irreducible to the separate and independent actions of parts of the mechanism over parts of the purview. This irreducibility is measured in terms of integrated information ϕ . Iterated at the system level, the bipartitions are applied to the set of all irreducible mechanisms and their purviews, which together form the systems’s ‘conceptual structure’. The minimal distance between intact conceptual structure and the conceptual structure under the partition is defined as the *integrated information* (denoted by Φ) of the system. A system is integrated, even if the least disrupting bipartition of the system, which is called the ‘minimum information partition’ (MIP) and partitions the system into two disconnected halves, would imply a loss of information in the causal power of the system (the partition is unidirectional and thus only cuts the connections from one half to the other but not the other way around). Instead of assessing whether a system is unified into a coherent whole by analysing its behaviour under regular conditions, IIT proposes that the forces integrating the behaviour of the system are better captured by observing its behaviour under perturbations. Integrated information formalises the interventionist notion of characterising the causal influences among the components of the

system [62, 63]. This formalism is believed to better capture the intrinsic causal structure of the system.

Recent studies have suggested that measures based on the general ideas of IIT can capture various states of lost consciousness, such as dreamless sleep [64], general anaesthesia [65], or vegetative states [66]. Some studies have suggested that integrated information could act as an order parameter of complex systems, similar to the generalised Ising model [67], coupled oscillators [68], and coupled mapping [69]. The magnitude of IIT 2.0 integrated information and the susceptibility of IIT 3.0 integrated information $\sigma(\Phi)$ peaks at critical points.

IIT has several versions [58–61, 70–72, 73] (Computational comparison of these versions of IIT was made by Mediano [74]). For example, there is a fundamental difference between IIT 2.0 [58, 59] and 3.0 [61]: one uses Riemann geometry and the other uses Wasserstein geometry. While both IIT 2.0 and 3.0 can capture phase transitions of the system, the differences occur in the behaviour at the criticality: Φ in IIT 3.0 does not necessarily peak at the critical point (the susceptibility of Φ peaks at the critical point) whereas φ in IIT 2.0 works as an order parameter. Note that Ito [75, 76] has pointed out that there are some intimate relations between the second law of information thermodynamics and IIT in terms of a projection onto a local reversible manifold. These structural resemblances suggest the possibility of unifying the concept of non-equilibrium thermodynamics and IIT 2.0. The extension of this idea to IIT 3.0 is still unknown.

Integrated information theory 3.0 is also resonant with complex systems because it posits that the whole cannot be reduced into its parts; the components produce synergetic information that would be lost upon separation [77, 78]. Complex biological systems are also irreducible owing to their intrinsic causal structures [79]. When applying IIT to these systems, Φ can be a measure of an autonomous system's wholeness [77], capturing intrinsic causal structures [80] while also acting as its order parameter. As an initial step to modelling collective animal behaviour, some researchers have used IIT to interpret classifications of cellular automata [54], animats [81], neuron-astrocyte ensembles [82], and Boolean networks [83]. For instance, the averaged max Φ values of major complexes for five to six automata cells correlated well with their complexity (e.g. class III and IV rules), despite the small cell-set number [54]. The behaviours of five and six automata cells are hard to discriminate on the basis of the behaviours of their constituent cells and, in general, the behaviours of small numbers of cellular automata are very similar with each other in terms of an external observer. They also showed that all rules of class IV have all orders of concepts (i.e. irreducible subsets in the system), unlike other classes. In addition to this study, many recent studies have found the connexion among IIT, group interaction, autonomy, criticality, edge of chaos, and a dynamical system (see [69, 84–87]). Linking IIT to collective behaviour seems very natural in this respect. Assuming that biological systems such as schools of fish evolved to reside in a critical state because of its advantages such as fast information transfer, the concept of network controllability using values of Φ may also provide meaningful insight into collective animal behaviour.

The example of cellular automata illuminates the meaning of intrinsic properties in IIT. IIT reveals the differences among systems arising from different intrinsic causal structures, rather than considering differences based on external behaviour. That is why we stated that previous approaches (e.g. TE) capture not 'what the system is' but 'what the system does'. This enables us to investigate the difference in collective behaviour in terms of the intrinsic causal structure. In this paper, we investigate the following: whether the number of agents in a system affects its intrinsic properties. In other words, if the group size changes, we would like to know what remains the same (continuous) and what changes (discontinuous) in the group. In addition, we consider whether there are any new factors introduced that were not present before. These kinds of questions are rarely asked in animal collective behaviour, but one study suggests that

schools of three fish and schools of two fish have different kinds of interactions [41, 42]. Another suggests that the search strategies of fish in groups of different sizes are essentially different when they are in an unfamiliar environment [11]. However, all these studies constrain the number of individuals in the group to three or less and their methods are difficult to generalise to larger groups. Furthermore, these methods never indicate any differences in terms of a group’s intrinsic causal structure.

In this paper, we apply IIT (in particular, IIT 3.0 using PyPhi [61, 72]) to schools of two to five fish (*Plecoglossus altivelis*) and show the intrinsic differences between these groups. We found that there is a kind of continuity and discontinuity with respect to school size. The main finding is that there is a discontinuity between three- and four-fish schools, which is a difference that has not received much attention previously. Interestingly, the difference between these two systems corresponds to the existence of leadership (more precisely, reducing the visual field for a fish’s recognition introduces the existence of the leadership). Furthermore, our results are never replicated by this discontinuity between three and four fish in a Boids-type model for the same conditions.

Results

Application of IIT 3.0 to fish school analysis

We tracked the trajectories of *ayu* fish schools of $N = 2$, $N = 4$, and $N = 5$ with three samples each, and $N = 3$ with four samples (10–15 minutes recording length; see [Materials and methods](#): Experimental settings for details). We then used these trajectories for analysis. To apply IIT 3.0, we define the ON and OFF states of an individual in a fish school. In this paper, an ON state means some interaction will occur in a given context. Note that in the Boids model [8, 88, 89], at least three kinds of interactions are needed: (i) interaction radius (distance), (ii) visual field, and (iii) turning rate (see Fig 1). (i) For interaction radius, if two individuals are within a

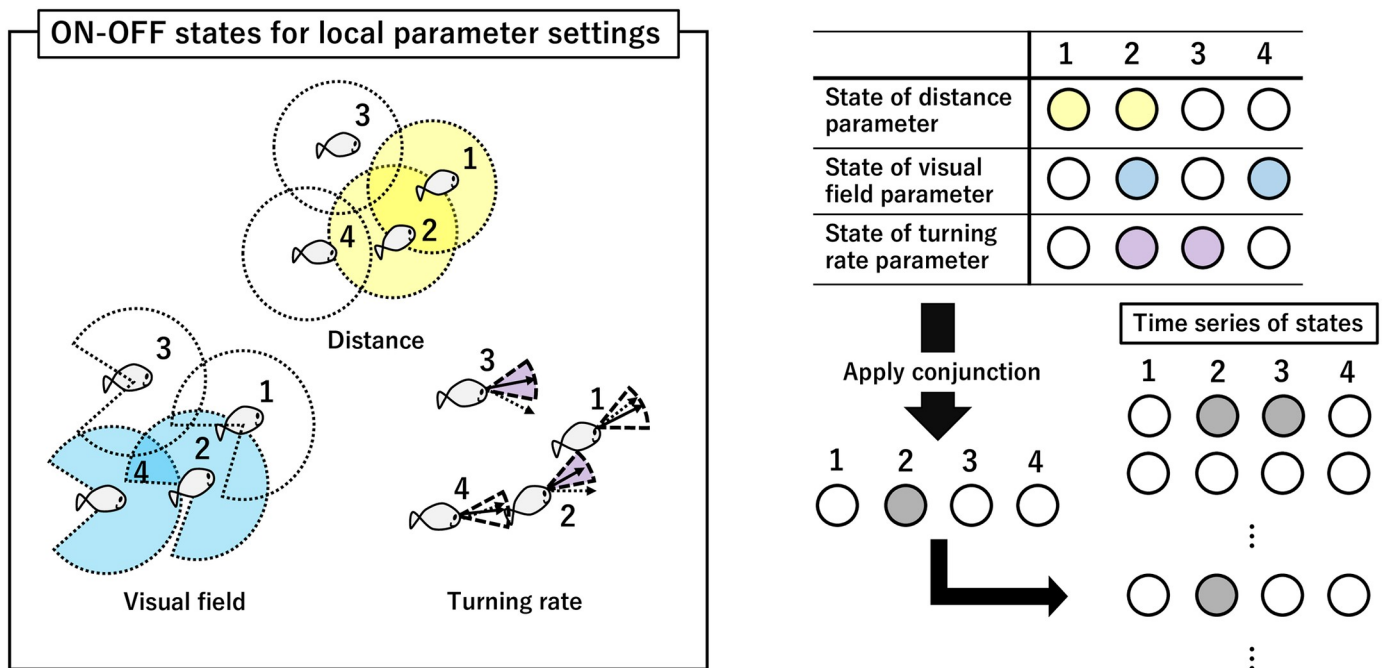


Fig 1. ON and OFF states for local parameter settings. Three parameters determine a school’s state (yellow: effective distance, blue: effective visual field, and purple: turning rate). Colour indicates the ON state.

<https://doi.org/10.1371/journal.pone.0229573.g001>

certain radius, the states of both individuals are ON (some information transfer will occur between them). This is a symmetric distance interaction. (ii) A visual field interaction means the individual is in the ON state when some other agents are within its visual field. This allows us to consider asymmetrical relations (in contrast to the symmetric distance condition). (iii) The turning rate interaction is one in which a direction change above a certain angle puts the individual in the ON state. This ON state transfers information to other agents in the next time step, so the interaction between individuals is a delayed one. The rate of direction change is a very important measure for collective behaviour, empirically and theoretically [8, 88, 89].

The first two parameters ('effect distance' and 'effective visual field' interactions) represent the interaction radius with the agent's blind spot (in later sections, we often refer to the environment outside of the visual field as the blind spot). The third parameter 'turning rate' represents the result of that interaction. Therefore, the final state is the conjunction of the states of all three parameters: distance, visual field, and turning rate (as detailed below). This representation reflects the commonly used interactions in a flocking model.

Note that the central interest of this paper is not the fixed individual's perceptual abilities (e.g. the fixed visual field) but the individual's effective perceptions under the dynamical interactions [13]. To state this more precisely, the interaction radius and the visual field should be interpreted as the 'effective' interaction radius and the 'effective' visual field.

To determine each fish state, the continuous trajectories of fish schools are quantised into 'Ising-like' binary time series. Biologically, such discretisation makes sense in terms of individual's contact, corresponding to the ON/OFF state, which is a good measure for a fish's communication in the school. Murakami et al. suggest that each fish's Levy walk behaviour in a *Plecoglossus altivelis* school enhances the efficient group communication through mutual contacts [10]. Therefore, it is reasonable to assume mutual contacts in the group reflect biological meaning to the group.

In this paper, we assume a fish always evaluates these three parameters simultaneously. Hence, we take their conjunction (i.e. $\text{AND}:\{0, 1\}^3 \rightarrow \{0, 1\}$). For the triple (1, 1, 1), $\text{AND}(1, 1, 1)$ is 1; otherwise, it is 0 to produce the overall state for a fish. For instance, IF (Distance = ON, Visual field = ON, Turning rate = OFF), THEN state = OFF. Note that the use of the AND operations reflects a Boid-like interaction (interaction radius with blind spot and direction changing). If any AND is replaced with an OR, the Boids-like interaction settings become meaningless. Applying the same process to each fish at time t , we obtain the time series of the states of the n fish. Then, we can compute Φ and other values (the sum of φ) from the obtained time series. One time step, in this paper, is defined as 0.05 to 0.10 s. This value roughly corresponds to a fish's reaction timescale [45].

To compute Φ , we also define the network structure in the school. In this paper, we postulate a completely connected network including self-loops (i.e. including self history). This assumption comes from the experimental fact that each fish has some contact with (or falls within the visual field of) all individuals in the group during a long series of recorded events (10–15 min). Therefore, it is natural to assume that some interactions have happened among all members. (In Table 3, we give the minimal distance that occurs during the events. The data show that all fish have a contact that is within 5 mm).

Before elaborating on our analysis, it is necessary to understand what the states ON and OFF mean for the fish. Biological information systems, such as the brain, have an explicit ON state, that is, when neurons are firing. In contrast, the ON state for each fish occurs when it interacts with its environment. Because there are various kinds of information to take into account, there is no explicit ON state for a fish school. Therefore, we computed all Φ values for all combinations of our parameter settings. Note that these combinations also contain only one or two of the three parameters. For instance, 2π rad given as the visual field and 0 rad/s

given as the turning rate change any individual state to ON. The distance parameter alone determines individual ON/OFF states in this condition (mathematically, this is $\text{AND}(x, 1, 1) = x$, where x is ON or OFF according to the distance parameter).

Φ values for real fish school data

MI, TE, and IIT 3.0. Mutual information (MI) and TE are often used to analyse dynamical systems. In this section, we explain how the Φ values of IIT 3.0 make it a unique measure compared with other well-known measures.

We compare three measures, that is, MI, TE, and Φ (IIT 3.0) under the same conditions. This comparison is possible because they share the same time series of an individual’s states (Fig 1). From these generated state sequences, we can compute the unique MI, TE, and Φ. In this paper, all analysis is restricted to the pair of parameters distance and visual field because the peak values always exist under a fixed turning rate condition of 0 rad/s. The importance of turning rate can be found over a longer time scale (see [90]). Instead, to verify that the following results are robust, we listed other time scale settings and Φ values as the main complex condition, adding to the results of other parameter settings, in the supporting figures.

Let the present and past states of the system at time t be given by $X(t) = \{x_1(t), x_2(t), \dots, x_n(t)\}$ and $X(t - 1) = \{x_1(t - 1), x_2(t - 1), \dots, x_n(t - 1)\}$. Elements $x_i(t)$ and $x_i(t - 1)$ respectively represent fish i ’s present and past states (and are 0 or 1), and n is the size of the fish school. Then, we have probability distribution $p(x_i) = \text{Pr}\{X = x_i\}$ and Shannon entropy $H(X) = -\sum_{i=1}^N p(x_i)\log(p(x_i))$. MI is expressed as follows:

$$MI = I(X(t); X(t - 1)) = H(X(t)) - H(X(t)|X(t - 1)) \tag{1}$$

This equation measures the shared information with the past (the states of the previous iteration, in this paper, $\Delta = 0.05$ s) and the present collective states. The collective state is the total ON and OFF sequence for the given parameter settings. For example, in the two fish schools, the collective state consists of four states, that is, $X = \{0, 0\}, \{0, 1\}, \{1, 0\},$ or $\{1, 1\}$.

In contrast, the TE measures the information transfer between two agents under the condition of knowing the other agent’s history (one step before the agent’s state is $\Delta = 0.05$ s). We define the TE as the sum of all pairwise transfer entropies.

$$TE = \sum_{\substack{(i,j) \in N \times N \\ i \neq j}} T_{x_i(t) \rightarrow x_j(t)} = \sum_{\substack{(i,j) \in N \times N \\ i \neq j}} (H(x_j(t)|x_j(t - 1)) - H(x_j(t)|x_i(t - 1), x_j(t - 1))) \tag{2}$$

The computation of IIT 3.0 is more complicated. For a given system in a particular state, all possible mechanisms (or subsets of system nodes in a state) that irreducibly constrain the past and future states of the system are identified. For each mechanism, all possible purviews or subsets of nodes that the mechanisms constrain are considered. For a given mechanism–perview combination, its cause–effect repertoire (CER; a probability distribution specifying how the mechanism causally constrains the past and future states of the purview). To find the irreducibility of the CER, the connexions between all permissible bipartitions of elements in the purview and the mechanism are cut; the bipartition producing the least difference is called the MIP. The irreducibility, or integrated information, ϕ , is quantified by the earth mover’s distance between the CER of the uncut mechanism and the CER of the mechanism partitioned by the MIP. A mechanism, together with the purview over which its CER is maximally irreducible and the associated ϕ value, specifies a concept, which expresses the causal role played by the mechanism within the system. The set of all concepts is called the cause–effect structure of the system. Once all irreducible mechanisms of a candidate system have been found, a similar set

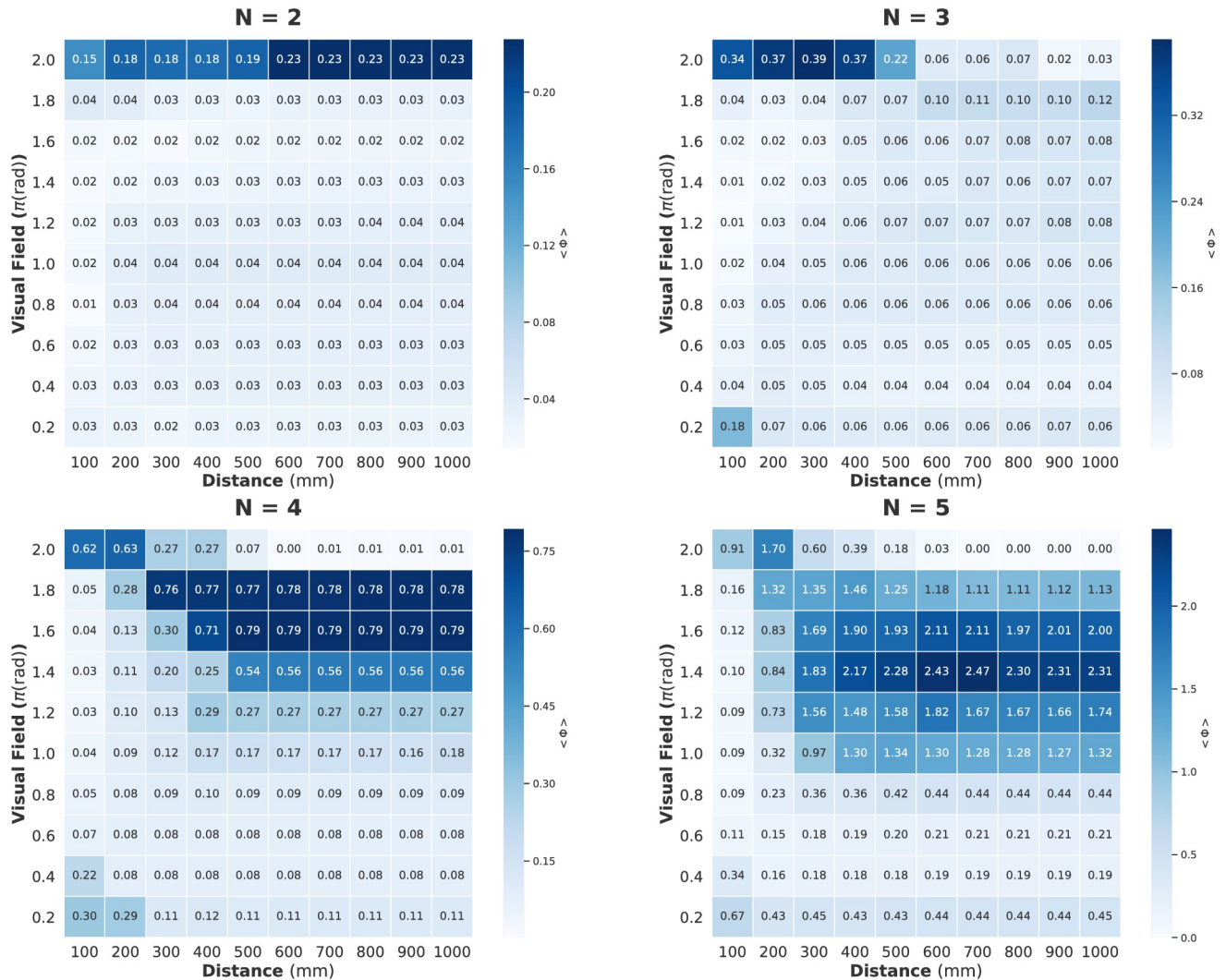


Fig 2. Heat maps showing the $\langle \Phi(N) \rangle$ values for distance (horizontal axis) and visual field (vertical axis) parameter values. Each cell value corresponds to the average $\langle \Phi(N) \rangle$ on all samples (three samples for $N = 2, 4, 5$ and four samples for $N = 4$). The value of $\langle \Phi(N) \rangle$ for two- and three-fish schools depends only on distance, whereas it depends on both parameters in four- and five-fish schools. The time step is $\Delta t = 0.05$ s.

<https://doi.org/10.1371/journal.pone.0229573.g002>

of operations is performed at ‘system level’ to understand at what level the set of mechanisms specified by the system are reducible to the mechanisms specified by its parts. The irreducibility of the candidate system is quantified by its conceptual integrated information Φ . This process is repeated for all candidate systems, and the candidate system that is maximally irreducible among all candidate systems is called the major complex.

We present a summary of IIT 3.0 in the Methods section. For readers unfamiliar with IIT 3.0, it is sufficient to interpret Φ as representing the degree of group integration. However, we have two remarks. (i) Each collective state has its own Φ . Hence, n -fish schools have a 2^n collective state. For example, in the case of a two-fish school in which all collective states are $(0, 0)$, $(1, 0)$, $(0, 1)$, and $(1, 1)$, we have four Φ s, that is, Φ_{00} , Φ_{10} , Φ_{01} , and Φ_{11} . (ii) The average $\langle \Phi(N) \rangle$ is that of all Φ values on all collective states, given by certain parameter settings. The values of the heat map in Fig 2 refer to the average of the mean $\langle \Phi \rangle$ from all data samples. In this paper,

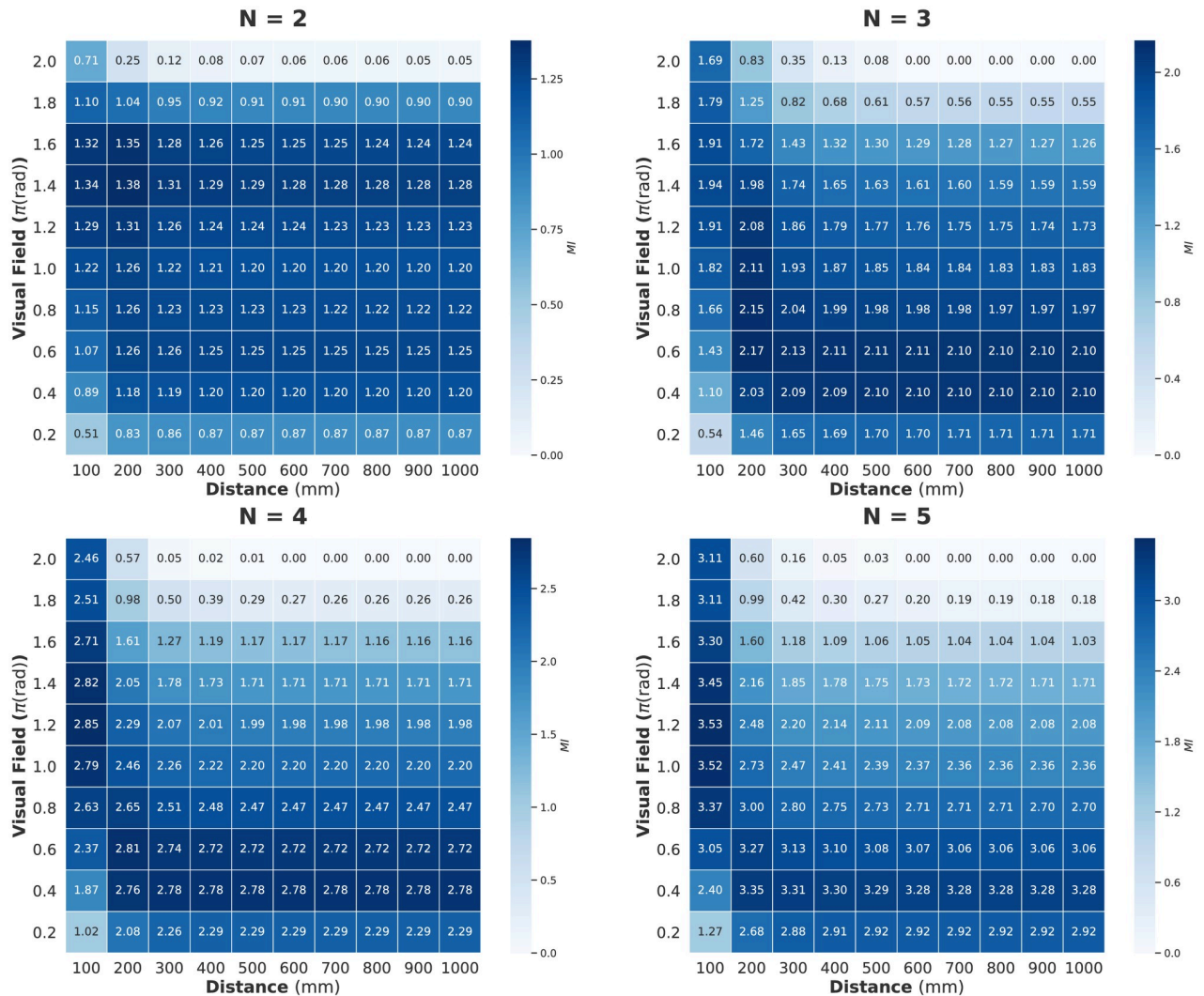


Fig 3. Heat map showing mutual information (MI) for distance (horizontal axis) and visual field (vertical axis) parameter values. Each cell value corresponds to the average MI on all samples. The time step is $\Delta t = 0.05$ s.

<https://doi.org/10.1371/journal.pone.0229573.g003>

although we computed n -size subsystem for n -fish schools, the distribution of Φ of the main complex also yields the same results [S4 Fig](#).

Figs [2](#), [3](#) and [4](#) show the computed values for two parameter settings (distance and visual field: turning rate is fixed at 0 rad/s); [Fig 2](#) shows $\langle \Phi(N) \rangle$, [Fig 3](#) shows MI, and [Fig 4](#) shows the sum of the TE. We can confirm the sharp contrast between $\langle \Phi(N) \rangle$ and the other two measures. The peaks of MI and TE are mainly concentrated in the low visual field area; whereas the peaks of $\langle \Phi(N) \rangle$ seem to have completely opposite trends (for the distribution including the turning rate, see [S1 Fig](#)). In this study, we mainly discuss the relation between these two parameter settings and $\langle \Phi(N) \rangle$ because the $\langle \Phi(N) \rangle$ values are not very high when the turning rate is non-zero). Interestingly, the distribution of $\langle \Sigma \varphi \rangle$ resembles that of MI and TE but not Φ ([S2 Fig](#)). The distribution of $\langle \Phi(N) \rangle$ remains the same when Δt is 0.1 s ([S3 Fig](#)).

A further outstanding fact is the existence of a discontinuity between three- and four-fish schools in the $\langle \Phi \rangle$ distributions. Note that the term ‘discontinuity’ (resp. ‘continuity’) that we use here is the inter-relation between two distributions for the group size and is not the intra-

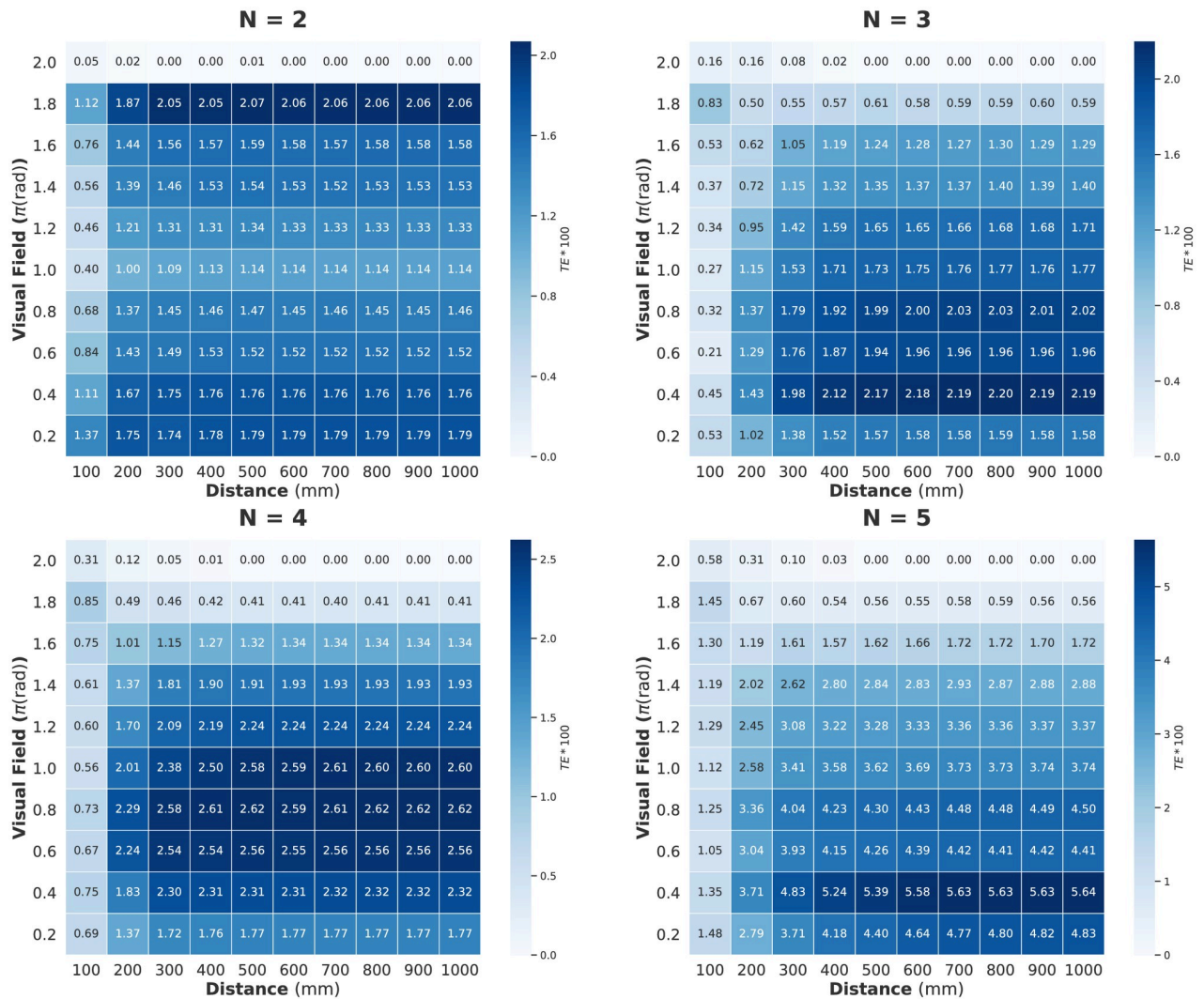


Fig 4. Heat maps showing the sum of transfer entropy (TE) for distance (horizontal axis) and visual field (vertical axis) parameter values. Each cell value corresponds to the average TE on all samples. Note that each TE value is multiplied by 100 for contrast). The distribution of TE resembles that of MI. In contrast, TE does not become larger than the observed MI. This is because each pairwise information transfer in large groups is less than those of small groups. The time step is $\Delta t = 0.05$ s.

<https://doi.org/10.1371/journal.pone.0229573.g004>

relation for the parameter settings in the same heat map. We can confirm that a qualitative shift happens between three- and four-fish schools (see Table 1). This trend was not observed in MI and TE (and $\langle \Sigma \phi \rangle$). All these facts imply that the $\langle \Phi(N) \rangle$ in IIT 3.0 captures something unique that is not observed when the traditional MI and TE metrics are used.

Table 1. Statistical test for the discontinuity on Fig 2. $\langle \Phi(N) \rangle_x$: means $\langle \Phi(N) \rangle$ from D = 100 mm to D = 1000 mm at fixed VF = x rad, where D is distance parameter and VF is visual field parameter. The p-values for $\langle \Phi(N) \rangle_{1.8\pi}$ and $\langle \Phi(N) \rangle_{2.0\pi}$ were computed using the Welch t-test. Values in bold correspond to the large mean Φ values. The opposite trend of $\langle \Phi(N) \rangle$ emerge between three- and four-fish school.

	$\langle \Phi(N) \rangle_{1.8\pi}$	$\langle \Phi(N) \rangle_{2.0\pi}$	p-value
N = 2	0.002	0.176	$< 10^{-10}$
N = 3	0.030	0.112	$< 10^{-7}$
N = 4	0.517	0.124	$< 10^{-8}$
N = 5	1.470	0.391	$< 10^{-8}$

<https://doi.org/10.1371/journal.pone.0229573.t001>

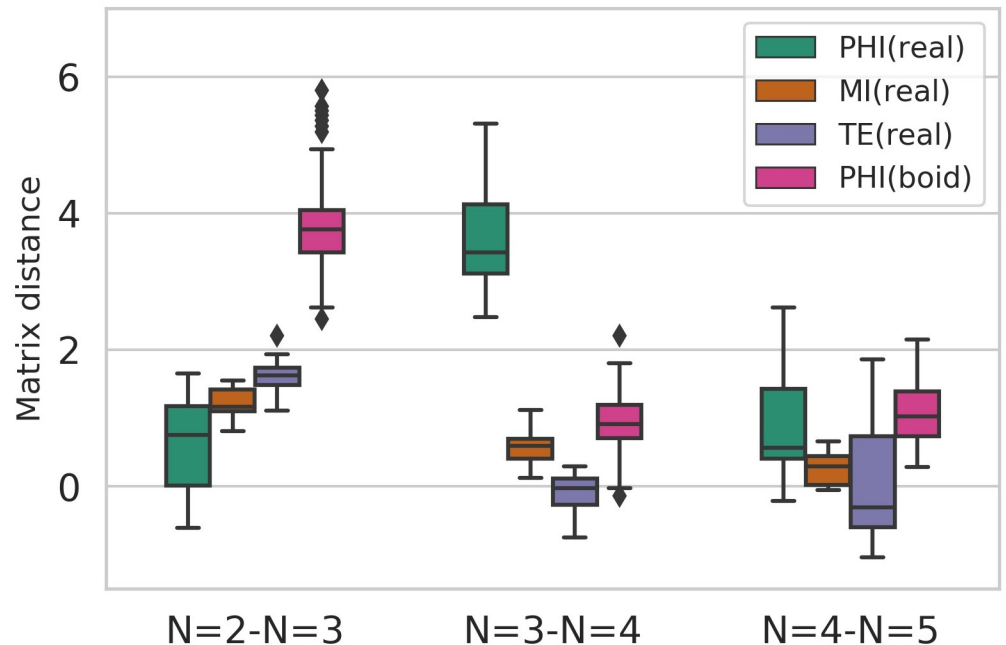


Fig 5. The mean matrix distance (MD) between inter-distributions subtracted by the baseline (the mean matrix distance among intra-distribution). The mean MD of PHI (real) between three- and four-fish schools is significantly larger than the other two measures and PHI (boid). PHI (boid) data is the same as the data shown in Fig 9.

<https://doi.org/10.1371/journal.pone.0229573.g005>

Fig 5 provides a whole picture on the difference between the inter-relation of each heat map (PHI (real), MI, TE, PHI (boid)): The data of PHI (boid) are in Fig 9). The vertical axis is the mean matrix distance (MD) between the inter-distributions of different group sizes. Note that in this study, we subtracted the MD from the baseline MD. This baseline MD corresponds to the average of all intra-distribution (N = 2, N = 3, N = 4, and N = 5) matrix distances for each measure.

The mean MD of PHI (real) between N = 3 and N = 4 distribution is much higher than the other two measures (Welch t-test: MI($p < 10^{-10}$), TE($p < 10^{-13}$) and PHI(boid)($p < 10^{-35}$). Furthermore, the mean MD of PHI (real) between N = 3 and N = 4 distributions is also higher than that between N = 2 and N = 3 (Table 1). In contrast, the other three show opposite trends, as illustrated in Fig 5. They have relatively large values for N = 2 and N = 3 than those of N = 3 and N = 4. Therefore, the graph suggests a discontinuity between three- and four-school distributions that were only confirmed in the $\langle \Phi(N) \rangle$ value distributions of real fish.

Emergence of leadership: Discontinuity between real three- and four-fish schools. We consider Fig 2 in detail. The concept of IIT 3.0 implies that Φ represents the group integration. If the group integration assigns some merit for group behaviour, to verify the meaning of high integration, it would be helpful to understand the collective behaviour. As we have already noted, there is a discontinuity between three- and four-fish schools. The striking contrast seems to exist when high visual field area ranges from 1.6π to 2.0π rad. Here, 2.0π rad means there is no blind spot for each fish; whereas the values less than 2.0π rad mean that each fish has some blind spot. Taking into account this contrast, fish schools of less than four show a high group integration (high $\langle \Phi(N) \rangle$) when their effective visual field is intact; in contrast, fish schools of four or more fish exhibit high group integration (high average $\langle \Phi(N) \rangle$) when their effective visual field has a blind spot.

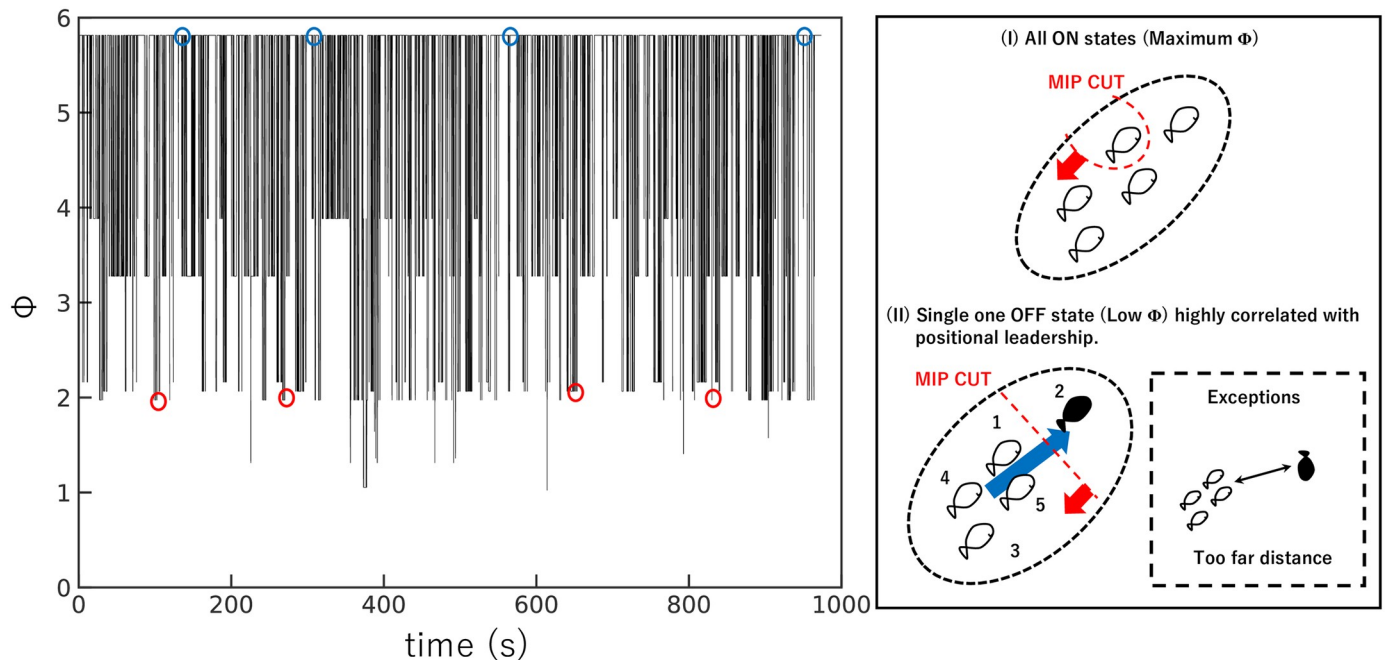


Fig 6. Example of a time series of Φ obtained from real fish data ($N = 5$). The reduction of Φ at various points indicates where leadership emerges in the group. The local parameter settings are distance = 700 mm, visual field = 1.6π rad, and turning rate = 0 rad/s, respectively. The illustrations on the right represent the example state of the fish (white: ON, black: OFF) and their minimum information partition (MIP). The blue circles on the left plot correspond to formation I on the right. The red circles on the left figure correspond to formation II on the right. This figure represents the cases where the positional leader (index 2) matches a single OFF state individual. The thick blue arrow is the average direction of the group. The dashed square indicates an exception, where a single OFF state individual is not the positional leader (this situation is rare: see Table 3). To enable interested readers to compute Φ using PyPhi, we list the transition probability matrix (TPM) in S1 Table.

<https://doi.org/10.1371/journal.pone.0229573.g006>

In this section, we explore this issue more deeply in terms of IIT. Focusing on whether the visual field parameter is 2π or not, we see the discontinuity on $\langle \Phi(N) \rangle$ distribution as the ‘emergence of leadership’. In this paper, we assume that the leader of the group should satisfy the following properties: (i) the leader is the head position of the group (i.e. the positional leader) and (ii) there is asymmetrical information flow among the leader and the rest of members. These kinds of properties seem ubiquitous in animal collective behaviour [91–94]. We aim to show a single OFF state in the group that highly correlates with these two properties, that is, (i) a single OFF state individual that shows high correlation with the positional leader and (ii) an MIP cut between the OFF-state fish and the rest of the ON-state fish corresponds to asymmetrical information flow in the group.

Fig 6 is a representative example that summarises our descriptions. Fig 6 is a time series of Φ when the visual field is 1.6π and the distance is 700 mm. Intermittent dropping Φ values are observed. These minima correspond to instances when there is a single OFF state in the collective state, that is, 01111, 10111, and so on. In contrast, the highest Φ value corresponds to the all-ON collective state, that is, 11111. Although there are some exceptions for different parameter settings or data samples, all MIP cuts are locate between the OFF-state fish and the rest of the ON-state fish.

First, we show that a single OFF-state fish highly correlates with positional leadership (property (i)). In this study, the positional leader is defined as the individual in the first position with respect to the average group direction (see Fig 6). We can compute the rate at which single OFF-state individuals actually correspond with the positional leader. Let N and T be sets $\{1, 2, \dots, N\}$ and $\{1, 2, \dots, T_{max}\}$, respectively, where T_{max} is the maximum time for a given

time step. Then, the single OFF state function is $OFF_{single}(t): T \rightarrow N \cup \{0\}$. This function returns the index of the OFF-state individual when the collective state has only one OFF state; otherwise, it returns 0. The positional leadership function is $P_{Leader}(t): T \rightarrow N$. This function selects the index of the positional leader.

$$MatchingRate(\%) = \frac{|\{t | OFF_{single}(t) = P_{Leader}(t), t \in T\}|}{|\{t | OFF_{single}(t) \in N, t \in T\}|} \times 100 \tag{3}$$

The above equation (Eq 3) indicates the frequency of occurrence of the OFF state individual as the positional leader. Table 2 shows that the matching rate increases substantially when the fish has a blind spot. If the fish has a maximum visual field, a single OFF-state individual does not correspond to the positional leader. In this context, it is appropriate to state that a positional leader emerges to raise the group integration on average when the group size is four or higher (an example of the actual positions for which single-OFF collective states occur are shown in S5 Fig).

Next, we examine the information flow between a single OFF-state individual and the rest of ON-state group members in terms of MIP (property (ii)). According to the depictions in Figs 6 and 7 shows the matching rate for the MIP cut and a single OFF-state individual under almost all parameter settings (in particular, high-distance regions in Fig 7 show perfect matches). The MIP cut in the groups in Fig 7 can be expressed in PyPhi as $\{1\} \not\Rightarrow \{2, 3, 4, 5\}$ or $\{2, 3, 4, 5\} \not\Rightarrow \{1\}$, where 1 is the index of a single OFF-state individual. The high correspondence of a single OFF-state individual and MIP cut locations suggests that the weakest link exists between the single OFF-state individual and the rest of the members (recall that the MIP cut lies where it causes the minimum information loss).

To summarise our analysis, the single-OFF state individual satisfies two properties that we have listed: (i) the leader in the head position of the group as the positional leader (Table 2), (ii) there is MIP induced asymmetrical information flow among the leader and the rest of the members. The single-OFF state in high $\langle \Phi \rangle_S$ condition for four- and five-fish schools becomes the leader in terms of IIT (hereafter, we call this single OFF-state individual as IIT leadership). Note that this asymmetrical information flow represents a ‘feedforward’ information flow to

Table 2. Φ values and match rates. The turning rate is fixed at 0 rad/s. All right-side values are average distance parameter values for all samples and range from 300 mm to 1000 mm. VF: visual field parameter; $\langle \Phi(N) \rangle_S$: mean Φ for a given set S, which is the set of single-OFF collective states such as {0111, 1011, 1101, 1110}; $\Phi_{11\dots 1}$: Φ values for the all-ON collective state. The p-values for $\langle \Phi(N) \rangle_S$ and $\Phi_{11\dots 1}$ were computed using the Wilcoxon signed-rank test. MR: match rate, computed by Eq (3) for a corresponding parameter set; $MR_{2\pi}$: match rate computed by Eq (3) for fixed VF = 2π . The p-values for MR and $MR_{2\pi}$ were computed using the Mann–Whitney U test. Values in bold correspond to the parameter settings for which the positional leader highly correlates with a single OFF-state fish in the group and where this fish also yields lower Φ values than those of $\Phi_{11\dots 1}$.

	VF (rad)	$\langle \Phi(N) \rangle_S$	$\Phi(N)_{11\dots 1}$	p-value	MR (%)	$MR_{2\pi}$ (%)	p-value
N = 4	1.2 π	0.169	0.200	< 0.001	95	13	< 0.001
	1.4 π	0.421	0.470	> 0.1	96	13	< 0.001
	1.6 π	0.676	0.592	< 0.001	96	13	< 0.001
	1.8 π	0.979	1.656	<0.001	94	13	<0.001
	1.9 π	0.805	1.630	<0.001	83	13	<0.001
N = 5	1.2 π	1.539	1.607	> 0.1	94	21	< 0.001
	1.4 π	2.884	3.831	<0.001	96	21	<0.001
	1.6 π	2.918	4.586	<0.001	96	21	<0.001
	1.8 π	2.199	4.617	<0.001	94	21	<0.001
	1.9 π	1.614	3.446	<0.001	82	21	<0.001

<https://doi.org/10.1371/journal.pone.0229573.t002>

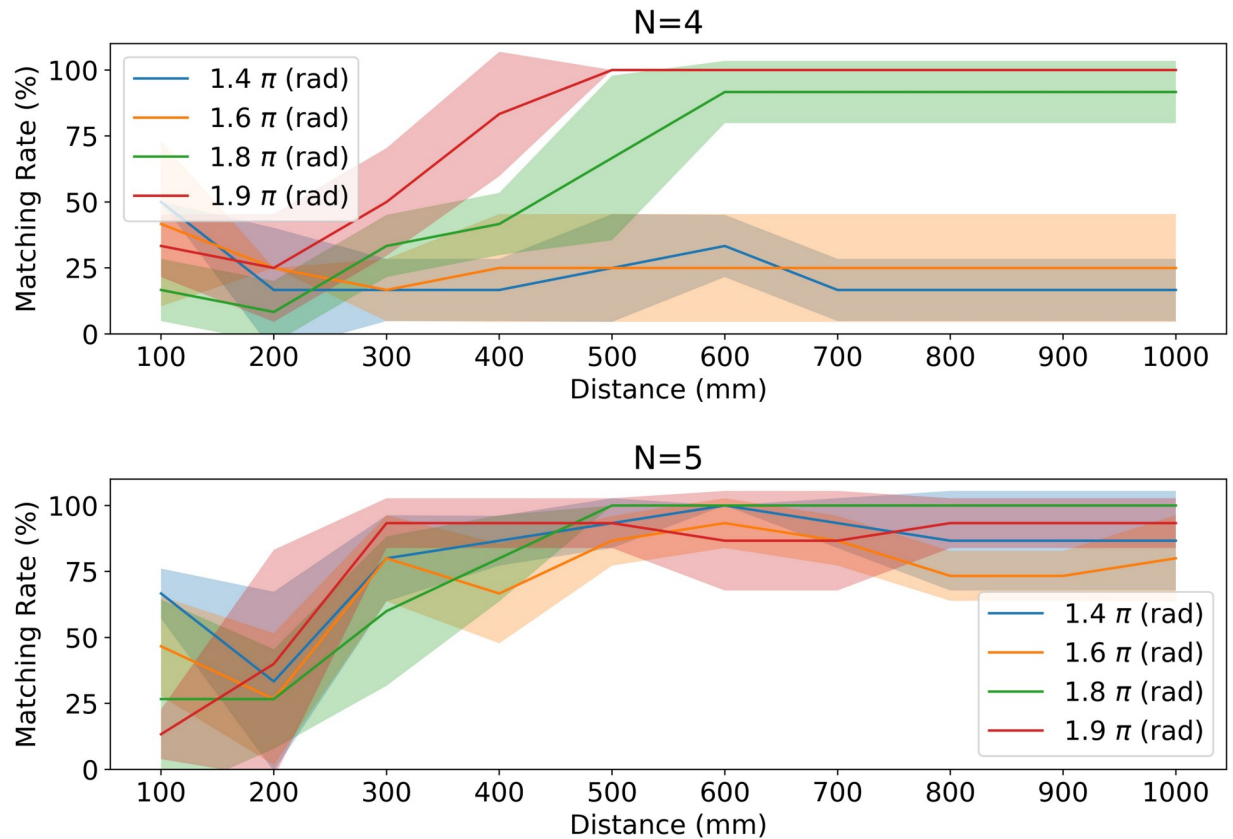


Fig 7. Match rates for a single OFF-state individual and the corresponding MIP cut between a single OFF state and all other ON states. We have $\{1\} \neq \{2, 3, 4, 5\}$ or $\{2, 3, 4, 5\} \neq \{1\}$ for the MIP cut when the collective state is 01111. The graph shows that the long-distance parameter discriminates the single OFF-state individual well according to the MIP cut for each visual field condition. Although a consistently low match rate exists for $N = 4$, these low visual field parameters are excluded by the leadership condition in Table 3. Note both graphs show an MIP cut without any self-loops because omitting them more clearly divides a single OFF-state individual and the rest of the members by an MIP cut (we have the same distribution as Fig 2 in S4 Fig). The version with self-loops is also listed in S6 Fig.

<https://doi.org/10.1371/journal.pone.0229573.g007>

Table 3. Data summary.

N	Data index	Average distance (mm)	Average velocity (mm/s)	Error (S.D.)	Minimum distance (mm)	Total Time Steps
2	1	166.3	268.8	0.18	1.90	106,961
	2	90.67	271.68	0.23	0.10	99,431
	3	122.0	256.08	0.18	1.60	107,206
3	1	170.8	301.2	0.23	1.80	90,051
	2	159.1	343.2	0.14	1.83	83,654
	3	173.1	300.0	0.13	2.82	97,446
	4	132.0	240.0	0.19	1.67	93,931
4	1	164.3	270.72	0.14	1.18	106,327
	2	141.5	190.8	0.12	1.38	103,226
	3	114.9	148.56	0.38	1.83	98,126
5	1	143.8	259.92	0.28	0.79	102,895
	2	146.0	213.12	0.12	1.16	97,346
	3	143.7	259.2	0.28	1.44	92,116

S.D. = standard deviation.

<https://doi.org/10.1371/journal.pone.0229573.t003>

or from the IIT leadership [61] because we have an MIP cut in both directions ($\{1\} \not\Rightarrow \{2, 3, 4, 5\}$ or $\{2, 3, 4, 5\} \not\Rightarrow \{1\}$).

The different MIP cuts can be interpreted as follows. In the case of $\{1\} \not\Rightarrow \{2, 3, 4, 5\}$, the Φ value indicates that information flows from the rest of the members to a single OFF-state individual, but not so much from a single OFF-state individual to the others. In contrast, the reverse relation applies for $\{2, 3, 4, 5\} \not\Rightarrow \{1\}$. Taking into account the low Φ at a single OFF-state condition compared with all-ON state conditions (see Table 2), the information flow from or to a single OFF-state individual is weak compared with its inverse flow.

These facts lead us to consider that classification would be possible according to the two kinds of IIT leaderships using these asymmetrical information flows, i.e., one is passive leadership (the former instance, because most of the information flows from the other members) and the other is active leadership (the latter instance, because most of the information flows from the IIT leader). In the Discussion section, we provide a response to some anticipated objections to our interpretation.

Representative Φ values for all data samples. Finally, we examine the average $\langle \Phi(N) \rangle$ values shown in Fig 2 in detail. In Fig 8, the peak cells in Fig 2 and S5 Fig (i.e. high average and high variance conditions) from each group size are as follows: $(N = 2: D = 600 \text{ mm and } VF = 2.0\pi \text{ rad}, N = 3: D = 400 \text{ mm and } VF = 2.0\pi \text{ rad}, N = 4: D = 600 \text{ mm and } VF = 1.8\pi \text{ rad}, N = 5: D = 700 \text{ mm and } VF = 1.6\pi \text{ rad})$. Here, D is the interaction radius and VF is the visual field. The turning rate is fixed at 0 rad/s . When increasing $\langle \Phi(N) \rangle$, the maximum Φ also increases (recall that an n -fish school has 2^n Φ s with respect to its collective states). This maximum Φ corresponds to the all-ON collective states for $N = 4$ or 5 (one example for $N = 5$ is used in Fig 6). The susceptibility is a measure of fluctuation or the response of an extensive property (such as the order parameter) to a small external perturbation that introduces some variations in the intensive property. Some researchers suspect that the susceptibility of integrated information defined by the standard deviation (or variance) of Φ relates to the given system's autonomous properties [67, 85]. Intuitively, a high standard deviation (or variance) indicates that the system contains both low group integration states and high group integration states. We show the variance distribution in S7 Fig for the parameter settings in Fig 2. Here, we only point out that the peak of $\langle \Phi(N) \rangle$ and variance of Φ do not always agree with each other.

IIT 3.0 on the Boids model

For comparison, simulated trajectories based on the Boids model [88] were analysed in the same manner as the trajectories of the real fish school. The Boids model was developed by Reynolds [95] and its complex and realistic-looking behaviour of a group of agents as a whole is determined entirely by the local interactions of individual agent choices based on a set of simple rules: repulsion, alignment, and attraction. In this study, N agents with position vectors x_i and unit direction vectors v_i were simulated in continuous two-dimensional space ($3,000 \times 2,500$) (the size of the experimental fish tank). Time was discretised into t computational time steps with a regular spacing $\Delta t = 0.05$. When there are n_r agents in the neighbourhood of agent i , the following rules for repulsion, alignment, and attraction were applied to update the variables of the agents at each t :

$$d_r(t + \Delta t) = -C \sum_{j \neq i}^{n_r} \frac{r_{ij}(t)}{|r_{ij}(t)|} \tag{4}$$

$$d_o(t + \Delta t) = C \sum_{j=1}^{n_j} \frac{v_j(t)}{|v_j(t)|} \tag{5}$$

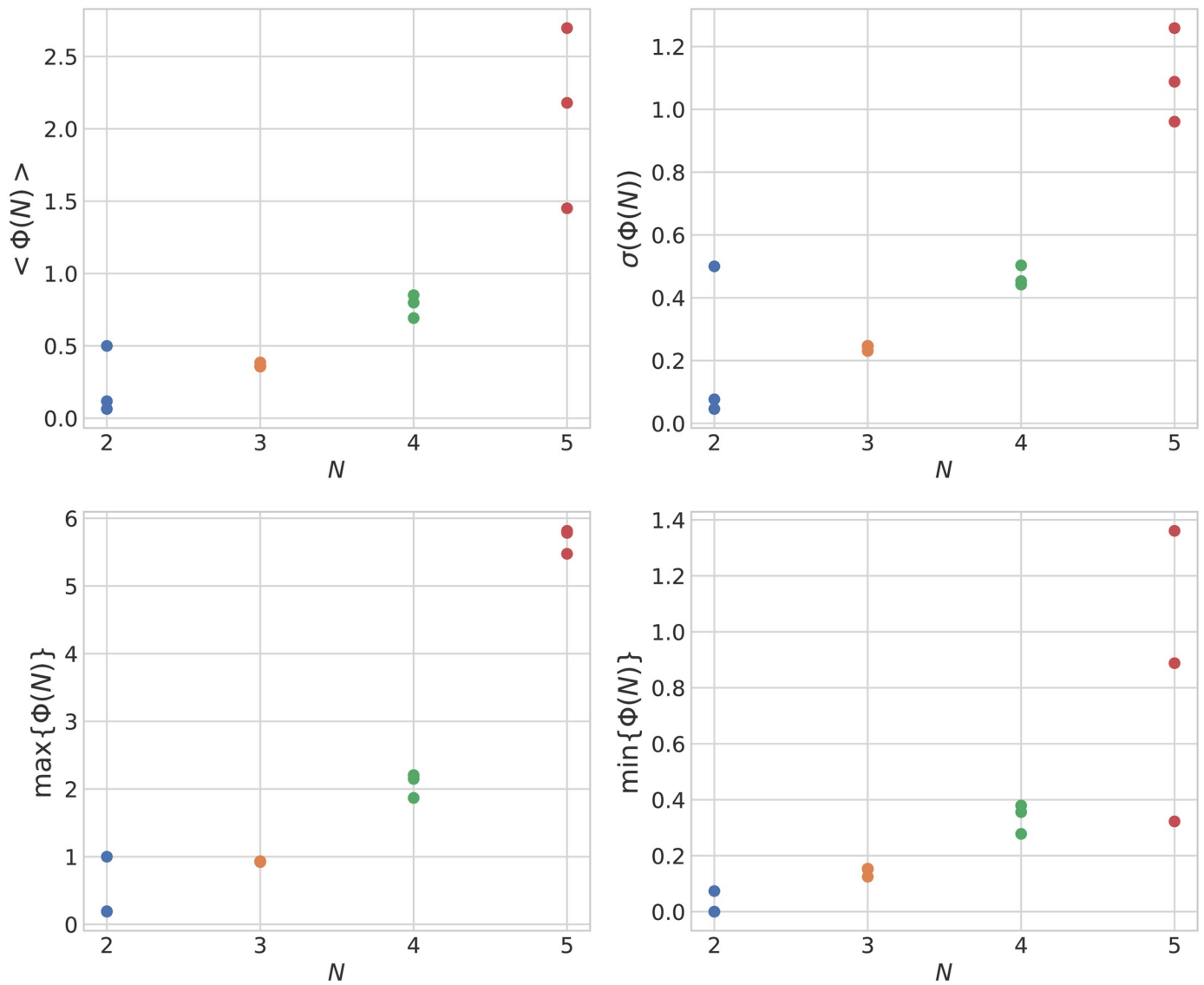


Fig 8. Mean values $\langle \Phi(N) \rangle$, standard deviations $\sigma(\Phi(N))$, max values $\max\{\Phi(N)\}$, and min values $\min\{\Phi(N)\}$ for various fish group sizes. Groups $N = 2$, $N = 4$, and $N = 5$ have three samples and group $N = 3$ has four samples. Parameters were set using the peak values of Fig 2. The mean values and standard deviations of integrated information Φ increase as the size of the fish school (N) increases.

<https://doi.org/10.1371/journal.pone.0229573.g008>

$$d_a(t + \Delta t) = C \sum_{i \neq j}^{n_a} \frac{\mathbf{r}_{ij}(t)}{|\mathbf{r}_{ij}(t)|} \tag{6}$$

where $n_r = \{j \mid \mathbf{r}_{ij}(t) \leq R\}$, $n_o = \{j \mid \mathbf{r}_{ij}(t) \leq O\}$, $n_a = \{j \mid O \leq \mathbf{r}_{ij}(t) \leq A\}$, and $\mathbf{r}_{ij} = \frac{(x_j - x_i)}{|(x_j - x_i)|}$ is the unit vector in the direction of neighbour j . The above rules were summed and averaged with additive Gaussian noise to determine the trajectories of the agents. The variables were updated synchronously. The parameters used in this simulation are shown in Table 4.

The trajectories of the Boids model and real fish showed the same complexity (S8 Fig); however, the $\langle \Phi \rangle$ heat maps for the Boids model exhibited very different patterns (Fig 9). The

Table 4. Summary of model parameters.

<i>N</i>	<i>R</i>	<i>O</i>	<i>A</i>	<i>v</i>	Error (S.D.)
2	10	120	∞	11.1	0.20
3	10	120	∞	12.4	0.17
4	10	120	∞	8.47	0.21
5	10	120	∞	10.2	0.23

S.D. = standard deviation.

<https://doi.org/10.1371/journal.pone.0229573.t004>

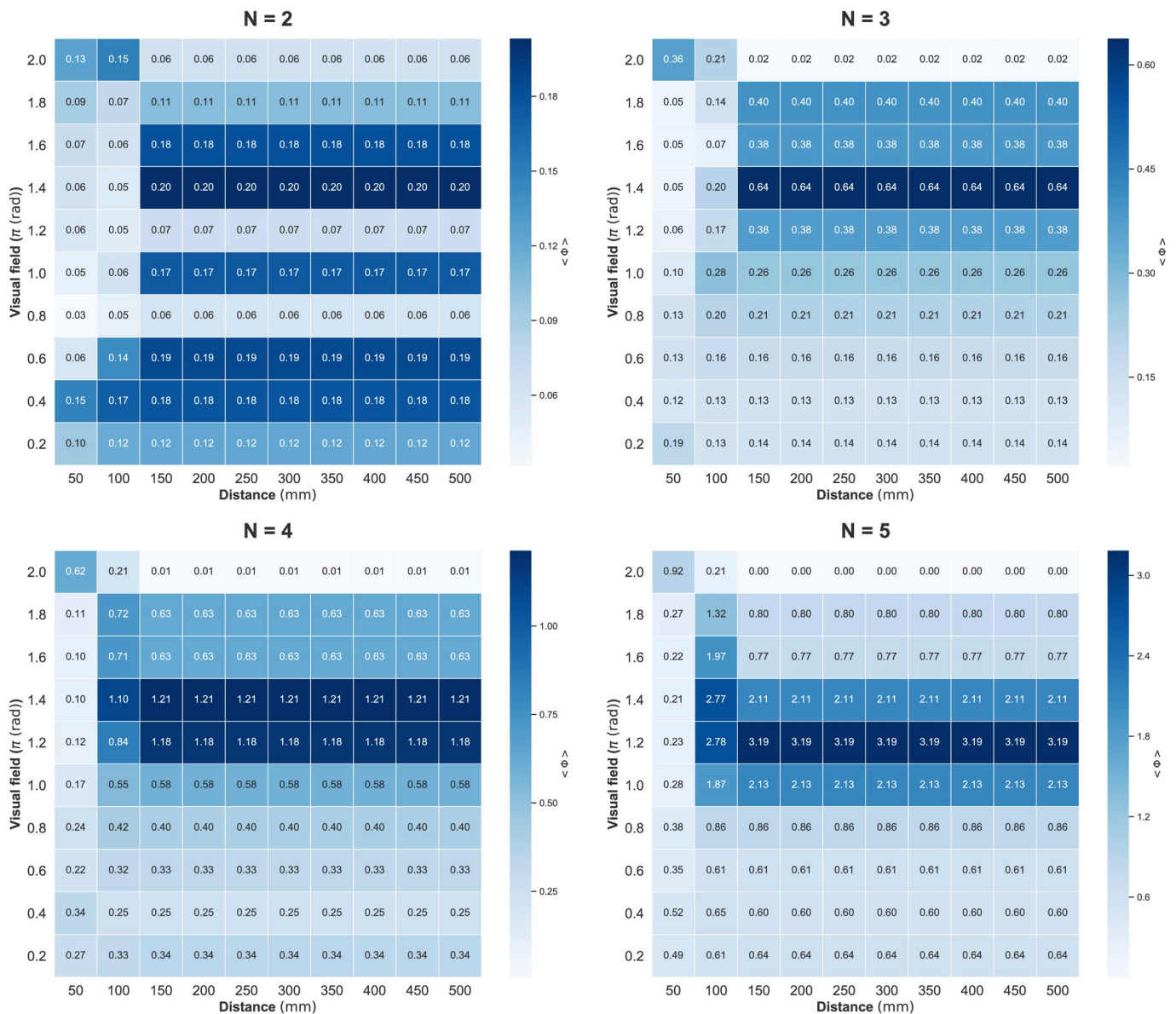


Fig 9. Values of mean $\langle \Phi \rangle$ for the Boids model for 10 samples each. Distance vs. visual field heat map.

<https://doi.org/10.1371/journal.pone.0229573.g009>

values of Φ for the Boids model were generally larger across the different parameter settings and did not seem to be as sensitive to the distance parameters.

The dynamics of the Boids model and real fish appeared to be very similar; however, the Φ values of the Boids model have large variances and lacked the discontinuity in their patterns between $N = 3$ and $N = 4$ (the distribution of MI and TE are resemble to real ones S9 Fig). In addition, there were substantial differences, especially when comparing the case of $N = 2$. When $N = 2$, the Boids model yielded wide distributions of Φ ; in contrast, the real $N = 2$ had very narrow and susceptible peaks.

The $\langle \Phi \rangle$ values were also calculated in the same manner and averaged for 10 simulations each. We were interested in any difference in the patterns of $\langle \Phi \rangle$ in the parameter space. We picked up the $\langle \Phi \rangle$ values for each of the following parameter configurations of the Boids model that maximise the $\langle \Phi \rangle$ values for the real fish school: Distance = 400 mm, Visual field = 2.0π rad for $N = 2$; Distance = 400 mm, Visual field = 2.0π rad for $N = 3$; Distance = 800 mm, Visual field = 1.8π rad for $N = 4$; and Distance = 700 mm, Visual field = 1.6π rad for $N = 5$. One parameter setting could maximise one $\langle \Phi \rangle$ value but not the other (Fig 10). The $\langle \Phi \rangle$ distributions of the Boids model were completely different from those of real fish.

To assess whether the maximisation of the integrated information is essentially dependent on the size of the system or fish behaviour, we investigated Boids models with reduced coupling strength ($C = 1.0, 0.1, 0.01, 0$) (see S10 and S11 Figs). We expected a coupling strength value of $C = 0$ to lead to a $\langle \Phi \rangle$ of 0 for all values of N ; however, $\langle \Phi \rangle$ is non-zero and increases as N grows. We suspect this is possibly due to size effects as well as boundary effects, both of which increase as N grows. Regarding the size effects, the Φ values of time-homogeneous Markov chains, which are probabilistic systems with random connexions, increase with size

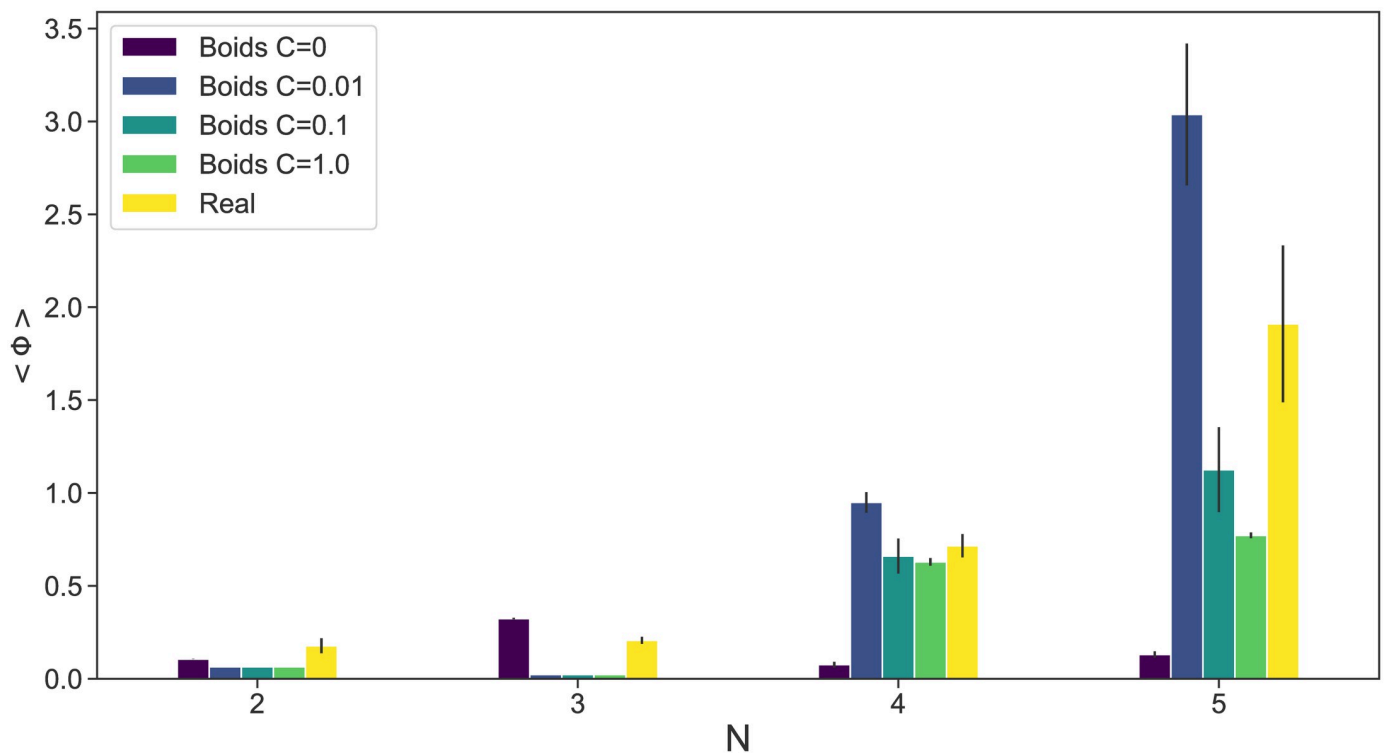


Fig 10. Comparison of the Φ values for certain parameter settings. The values of Φ_{Boids} were calculated at the parameter configurations that maximise the real fish $\langle \Phi \rangle_{\text{Real}}$ and averaged for 10 simulations. Error bars indicate the standard deviations.

<https://doi.org/10.1371/journal.pone.0229573.g010>

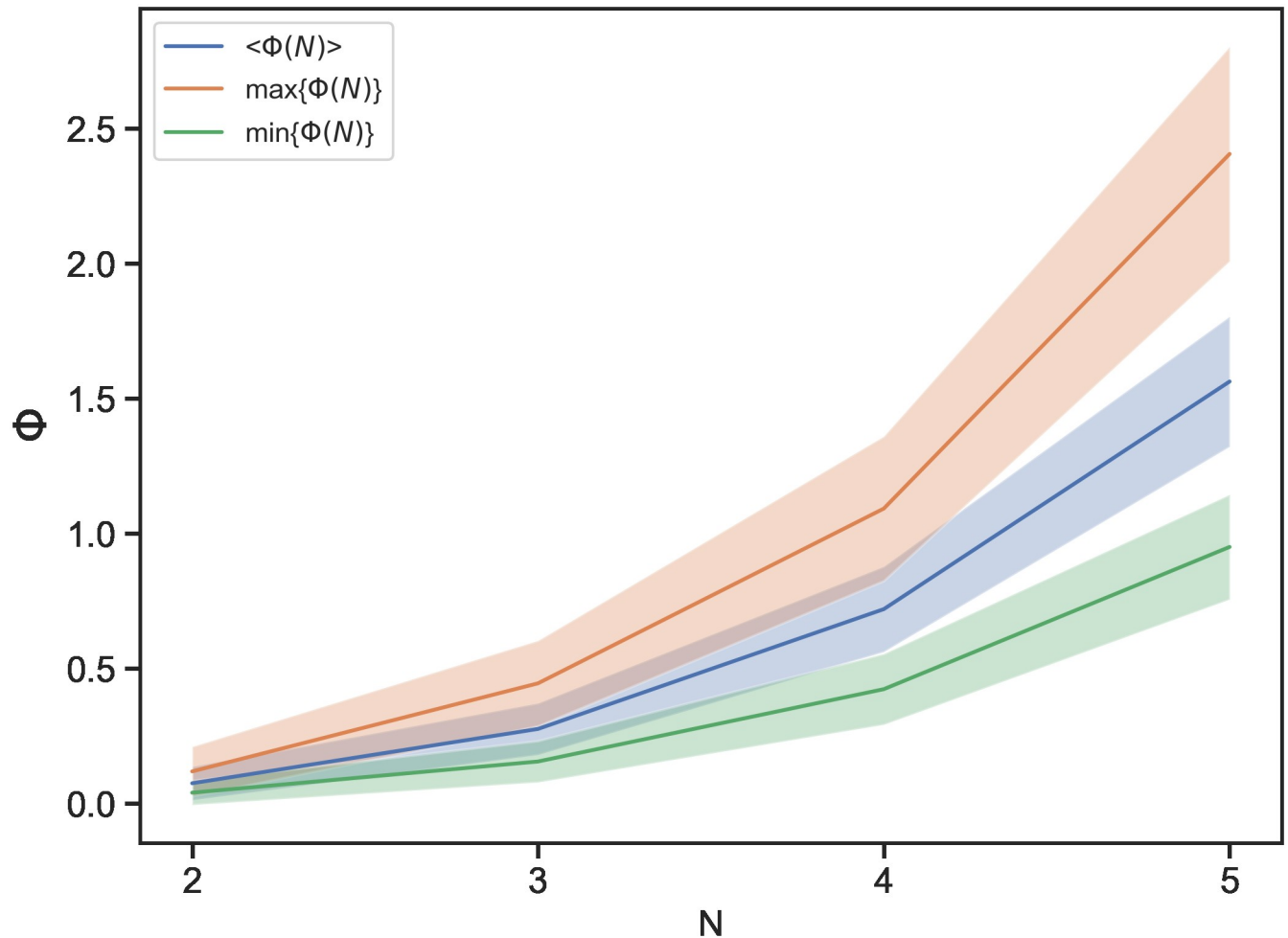


Fig 11. $\langle \Phi \rangle$, $\max\Phi(N)$, and $\min\Phi(N)$ values of time-homogeneous Markov chains with different values of N . The Φ values were calculated for randomly generated 100 transition probability matrices (TPMs). The shadows indicate the standard deviations.

<https://doi.org/10.1371/journal.pone.0229573.g011>

(Fig 11). Comparing the real fish schools and the Markov models, it is remarkable that the real fish schools have larger standard deviations for $\Phi(N)$ (Fig 8) than those of the Markov chains ($\sigma(\Phi(2)) = 0.03 \pm 0.04$, $\sigma(\Phi(3)) = 0.09 \pm 0.03$, $\sigma(\Phi(4)) = 0.19 \pm 0.06$, and $\sigma(\Phi(5)) = 0.36 \pm 0.07$). The larger variances of Φ in the real fish schools imply that the real fish have more variety and diversity in their states.

Another factor that raised the $\Phi(N)$ values for $C = 0$ systems was the use of the binary parameter format, even though the fish move continuously. The $\Phi(N)$ values would be 0 if the fish were placed randomly at all time steps so that there were no interactions (no boundaries). For example, for a visual field of 2π , if one fish is in the visual field of another, the reverse is also true; i.e., the state 1000 is impossible, only 1100 and similar patterns can exist. In such conditions, asymmetric interactions in which one fish is in the ON state and the other remains in the OFF state can never occur. This heterogeneity in TPM results in the non-zero Φ . We call this the boundary effect. Interestingly, the weak interactions $C = 0.01$ that have different distributions compared with $C = 0$ might suppress the boundary effects, resulting in a relatively lower $\langle \Phi \rangle$. In fact, the model with weaker interactions $C = 0.01$ actually has a distribution that is very similar to that of a real fish school S10 Fig). While the averaged $\langle \Phi \rangle$ depends on

coupling strength, we emphasise that we must compare the distributions to properly discuss the effect of behaviour on integrated information [S11 Fig](#)). A clear change in the distributions of $\langle \Phi \rangle$ occurs when coupling strength is reduced.

Discussion

Although IIT was first proposed to quantify the degree of consciousness from brain activity, interest in IIT seems to have gradually shifted to using it to determine a system's causal structure itself [54, 80, 83]. More precisely, the researchers' concern for the causal structure of the autonomous system was hidden under their ambitious project to quantify human consciousness. This move is commendable because, by doing so, IIT researchers have paved the way for a new measure for any general living system. The concept of causal structure is not used to determine information flow among agents, as in TE. Instead of measuring 'what the system does', IIT can measure 'what the system is'. By imposing an intervention on the system, we can estimate how the system reacts to such external perturbations. Pearl, known as the founder of causal inference, and other researchers have also applied almost the same perturbation method to causal networks [62]. The MIP cut makes integrated information stand apart from other information measures like MI or TE [96].

In this study, we applied IIT 3.0 to real fish schools and compared the results with those for other measures (MI and TE) and another model (Boids) under the same conditions. The simple phenomenon that IIT suggests is an increase in $\langle \Phi(N) \rangle$ values with group size in real and Boids-model fish schools. Furthermore, the coupling strength of the Boids model also changes the $\langle \Phi(N) \rangle$ values. This result agrees well with our intuition for the group integration using IIT 3.0. Our experiment indicates that a relatively weak interaction happens in real fish schools.

From the Φ distributions for all possible combinations of parameters, we found a discontinuity between three- and four-fish schools: the leadership raises the degree of integration above four but not below three. This discontinuity between three and four fish not only exists in the IIT of real fish but also in the IIT of the Boids model. Furthermore, we found that MI and the sum of TE also fail to capture these differences. Note that IIT discriminates between three- and four-fish groups, which is a comparison that is rarely considered in the context of collective animal behaviour, although there are some studies that suggest a difference between two- and three-fish groups in terms of each fish's interactions with others (i.e. a difference in 'what the system does' from an extrinsic perspective) [11, 41, 42].

Finally, we comment on the IIT leadership in the group. The definition of IIT leadership is a single OFF-state in the collective state (e.g. 11110 or 01111) which satisfies the following: (i) matching well with positional leadership (the leader of the group is the head of its group) and (ii) the MIP cut is located between a single OFF-state individual and the rest of the group members.

Prior to providing details of leadership discussion, two objections might still be anticipated. One is that if the MIP cut indicates the weakest connection (interaction) among the leader and the rest of its members, does it mean that the low $\langle \Phi \rangle$ suggests a relatively independent relation between the leaders and the rest of the members? As discussed earlier, MIP cut is unidirectional and not bi-directional. The low $\langle \Phi \rangle$ values induced MIP provide no information about opposite causal flows are also weak. Adding to this point, the $\langle \Phi \rangle$ values of single OFF-state collective states (01111) in five-fish schools show higher values compared with all one states (1111) in four-fish school as in [Table 2](#). Although the group integrity is low when the five-fish schools are at a single OFF-state collective states, their integrity is still sufficiently higher than the highest integrity in four-fish school. Therefore, the first objection can be rejected.

The next objection is (ii) If the opposite information flows are as weak as the MIP cut ones (i.e. both flows have the same weakness), it is hard to suggest the feed-forward relation as we discussed. This objection is difficult to answer because PyPhi provides no information about opposite direction information flows. However, if this assumption is correct, we consider that it is not so critical that our interpretation is incorrect. This assumption is possible when we suppose that the weak interactions never contribute to the group dynamics. This assumption is not always true, especially in a dynamical system. For example, some researchers suggest that the weak interaction provides the stabilisation/destabilisation to the ecosystem [97–99]. Even if small perturbations are added to the system, the degree of perturbations can work as the system's parameters and make the system qualitatively different through bifurcation. Therefore, the weak interaction itself never suffices to spoil our interpretation of the leadership induced MIP.

The group leader detection result using IIT 3.0 suggests two remarkable propositions for understanding group behaviour. One is that our natural intuition of leadership (i.e. positional leadership) is basically correct under the given causal structure. As in [61], low Φ values represent that the system is a feed-forward system (more precisely, it has a weak recurrent interaction structure). In this sense, the information flow is uni-directional because it is in a low Φ s system. Therefore, the low Φ s of single-OFF collective states have much lower recurrent interactions than the all-ON collective states.

The lack of recurrent interactions also gives a new insight into the discontinuity between three- and four-fish schools. If the visual field is not 2π for two- and three-fish schools, their interactions are almost completely feed-forward in all collective states. The visual field asymmetry never contributes to recurrent interactions. As a result, the leadership is not explicitly discriminated by IIT 3.0; however, if the visual field is not 2π for fish schools with four or more fish, their interaction structure becomes more differentiated. Although Φ values are low under the single-OFF condition, the high Φ values of the all-ON condition suggest that a profoundly organised recurrent interaction emerges. Therefore, the discontinuity between three- and four-fish schools can be interpreted in terms of their different information structures.

The other proposition is that the MIP may detect the division of roles in collective behaviour (i.e. the weakest link divides the group into related aggregations). If we can extend this method to larger fish schools, IIT 3.0 would be able to determine the hidden division of the roles of a given collective behaviour. Furthermore, this role may not be revealed by other information measures.

In this study, we avoided going deeper into the problem of timescale (we only used a relatively small timescale, which is roughly equal to a general fish's reaction time). Over longer timescales, other patterns of continuity and discontinuity may be found. Increasing the number of individuals may also give other results. However, the present practical computational limit of IIT 3.0 is around seven or eight individuals or neurons [72]. Hence, some approximations will be needed to implement further analysis. Another area we did not address is the network structure. We assumed an all-connected network including self-loops in this study because all fish come into contact with each other throughout the event. As we hinted in S6 Fig, the network structures that exclude self-loops might be more suitable for finding the division in terms of MIP. Furthermore, some studies suggest that the network structure of real schools of fish is radically different from the Boids model and that they instead create a stable network called the α -lattice [100, 101]. This type of network may prevent the increases in Φ observed in the Boids model.

Materials and methods

Ethics statement

This study was carried out in strict accordance with the recommendations in the Guide for the Care and Use of Laboratory Animals of the National Institutes of Health. The protocol was approved by the Committee on the Ethics of Animal Experiments of the University of Tsukuba (Permit Number: 14-386). All efforts were made to minimize suffering.

Computation of Φ

All computations, in this paper, were performed using the PyPhi software package with the CUT_ONE_APPROXIMATION option for Φ .

Experimental settings

We studied *ayus* (*Plecoglossus altivelis*), also known as sweetfish, which live throughout and are widely farmed in Japan. Juvenile *ayus* (approximately 7–14 cm in body length) display typical schooling behaviour, though adult *ayus* tend to show territorial behaviour in environments where fish density is low. We purchased juveniles from Tarumiyoushoku (Kasumigaura, Ibaraki, Japan) and housed them in a controlled laboratory. Approximately 150 fish lived in a 0.8 m³ tank of continuously filtered and recycled fresh water with a temperature maintained at 16.4°C, and were fed commercial food pellets. Immediately before each experiment was conducted, randomly chosen fish were separated to form a school of each size and were moved to an experimental arena without pre-training. The experimental arena consisted of a 3×3m² shallow white tank. The water depth was approximately 15 cm so that schools would be approximately two-dimensional. The fish were recorded with an overhead grey-scale video camera (Library GE 60; Library Co. Ltd., Tokyo, Japan) at a spatial resolution of 640×480 pixels and a temporal resolution of 120 frames per second.

Data summary

The data are summarised in [Table 3](#).

Parameters are the key to determining the dynamics of the model. In the present study, the model parameters were set to simulate the real experimental data shown in [Table 3](#). The average distances were approximately 80 to 140 mm, so we set $O = 120$ mm, $R = 10$ mm (the body length), and $A = \infty$. Thus, the fish schools should not part less than 140 mm. This setting was necessary for the school to become separated by the boundary conditions, which mimic and reflect the real data. The amplitudes of the noise was set to be proportional to the average angle change so each agent would have a different level of noise. The parameter settings are shown in [Table 4](#).

Definition of ON and OFF states for each parameter

We define a function for each parameter that returns either 0 (OFF) or 1 (ON) for a given input value. Generally, we denote a function as $F_i^t(\cdot)$, where F is the name of the function, i is the index of the individual and t is the time. The arguments of the function can be either in the position vectors $\mathbf{x}_i(t)$ or the velocity vectors $\mathbf{v}_i(t)$ of each individual at time t . In general, the dimensions of these vectors are $d \leq 3$; the experimental setup used here gives $d = 2$. The number of individuals is n .

Parameter settings.

- Distance function $D_i^t(\mathbf{x}_1(t), \mathbf{x}_2(t), \dots, \mathbf{x}_n(t)) : \mathbb{R}^d \times \mathbb{R}^d \times \dots \times \mathbb{R}^d \rightarrow \{0, 1\}$
 For each individual i we obtain a set $S_i^t = \{j | d(\mathbf{x}_i(t), \mathbf{x}_j(t)) < \zeta, j \neq i\}$ of all other individuals within a specified distance ζ . Here $d(\mathbf{x}, \mathbf{y})$ gives the Euclidean distance between \mathbf{x} and \mathbf{y} .
 Then, $D_i^t(\mathbf{x}_1(t), \mathbf{x}_2(t), \dots, \mathbf{x}_n(t)) = 1$ when $|S_i^t| > 0$ and is 0 otherwise, where $|S|$ denotes the number of elements of a set S .
- Visual field function $B_i^t(\mathbf{x}_1(t), \mathbf{x}_2(t), \dots, \mathbf{x}_n(t), \mathbf{v}_1(t), \mathbf{v}_2(t), \dots, \mathbf{v}_n(t)) : \mathbb{R}^d \times \mathbb{R}^d \times \dots \times \mathbb{R}^d \rightarrow \{0, 1\}$
 For each individual we form the set $O_i^t = \{j | \arg(\mathbf{v}_i(t), \mathbf{x}_i(t) - \mathbf{x}_j(t)) < \eta, j \neq i\}$ of all other individuals whose velocity vectors point in a direction within an angle η of that of the focal individual. The function $\arg(\mathbf{x}_1(t), \mathbf{x}_2(t))$ gives the angle between two vectors. Then,
 $B_i^t(\mathbf{x}_1(t), \mathbf{x}_2(t), \dots, \mathbf{x}_n(t), \mathbf{v}_1(t), \mathbf{v}_2(t), \dots, \mathbf{v}_n(t)) = 1$ when $|O_i^t| > 0$ and is 0 otherwise.
- Turning rate function $T_i^t(\mathbf{v}_i(t), \mathbf{v}_i(t - \Delta t)) : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \{0, 1\}$
 The turning rate function returns 1 when an individual's turning rate exceeds a specified threshold δ . That is, $T_i^t(\mathbf{v}_i(t), \mathbf{v}_i(t - \Delta t)) = 1$ when $\arg(\mathbf{v}_i(t), \mathbf{v}_i(t - \Delta t)) \geq \delta$ and is 0 otherwise. The time step used in this paper is $\Delta t = 0.05$, $\Delta t = 0.1$ or $\Delta t = 0.2$ s.
 To obtain the states of the fish school, we take a conjunction of this result, that is, $D_i^t(\mathbf{x}_1(t), \mathbf{x}_2(t), \dots, \mathbf{x}_n(t)) \wedge B_i^t(\mathbf{v}_1(t), \mathbf{v}_2(t), \dots, \mathbf{v}_n(t)) \wedge T_i^t(\mathbf{v}_i(t), \mathbf{v}_i(t - \Delta t))$ for each individual i . The conjunction is given as $\wedge : \{0, 1\}^2 \rightarrow \{0, 1\}$ where $1 \wedge 1 = 1$ and is 0 otherwise. Thus the state of each individual i at time t is $s_i(t; \zeta, \eta, \delta) \in \{0, 1\}$ which depends on the triplet of parameter values (ζ, η, δ) . The state of the school at time t is then a vector $s(t) = (s_1(t), s_2(t), \dots, s_n(t)) \in \{0, 1\}^n$, where the parameter dependence has been omitted for simplicity.

Short summary of integrated information Φ

Integrated information theory models a system S by the discrete time multivariate stochastic process

$$p(X_0, X_{\Delta t}, \dots, X_t, X_{t+\Delta t}, \dots, X_T) \tag{7}$$

which fulfils the Markov property

$$p(X_0, X_{\Delta t}, \dots, X_t, X_{t+\Delta t}, \dots, X_T) = p(X_0) \prod_{i=\Delta t}^T p(X_i | X_{i-\Delta t}) \tag{8}$$

Such a discrete dynamical system S is defined by a directed graph of interconnected nodes (in this study, we assumed a complete graph.) and its TPM. The TPM specifies the conditional probability distribution $p(X_t | X_{t-\Delta t})$. Each state vector X_t comprises binary variables $x_i, i = 1, 2, \dots, n (n \in \mathbb{N})$.

A joint distribution $p_{\text{cause-effect}}$ is defined as

$$p_{\text{cause-effect}}(X_{t-\Delta t}, X_t) := p_u(X_{t-\Delta t}) p_{\text{effect}}(X_t | X_{t-\Delta t}) \tag{9}$$

The marginal distribution $p_u(X_{t-\Delta t})$ is a uniform distribution to give the maximum entropy distribution.

From the joint probability above, the backward transitional probability distribution

$$p_{\text{effect}}(X_{t-\Delta t} | X_t) := \frac{p_{\text{cause-effect}}(X_{t-\Delta t}, X_t)}{\sum_{X_{t-\Delta t}} p_{\text{cause-effect}}(X_{t-\Delta t}, X_t)} \tag{10}$$

and the forward transitional probability distribution

$$p_{\text{cause}}(X_t | X_{t-\Delta t}) := p((X_t | X_{t-\Delta t})) \tag{11}$$

are constructed and referred to as the *cause repertoire* and the *effect repertoire* of state X_t , respectively. The *cause repertoire* and the *effect repertoire* are calculated for a set of nodes within the subsystem, or a *mechanism* $M \subseteq S$, over another set of nodes within the subsystem, or a *purview* of the mechanism.

After assessing the information of a mechanism over a purview, we next consider its *integrated information* $\varphi_{\text{cause-effect}}$ of a set of system elements in a state X . The integrated information is defined as

$$\phi_{\text{cause-effect}} := \min\{\phi_{\text{effect}}, \phi_{\text{cause}}\} \tag{12}$$

$$\phi_{\text{effect}} := \min_{i \in I} \{D(p_{\text{effect}} \| p_{\text{effect}}^{(i)})\} \tag{13}$$

$$\phi_{\text{cause}} := \min_{i \in I} \{D(p_{\text{cause}} \| p_{\text{cause}}^{(i)})\} \tag{14}$$

where the system is decomposed in all possible ways into I .

The integrated information φ is assessed by quantifying the extent to which the cause and effect repertoires of the mechanism-purview pair can be reduced to the repertoires of its parts. The amount of irreducibility of a mechanism over a purview with respect to a partition is quantified as the divergence between the unpartitioned repertoire p and the partitioned repertoire $p^{(i)}$. The partition that yields the minimum irreducibility is called the *minimum-information partition* (MIP). The integrated information φ of a mechanism-purview pair is defined as the divergence between the unpartitioned repertoire and the repertoires partitioned by the MIP. The maximum φ value is then searched for over all possible purviews to determine the *maximally-irreducible cause* and *maximally-irreducible effect* specified by a mechanism.

$$\varphi_{\text{cause}}^{\max} := \max_{j \in C} \{\varphi_{\text{cause}}^j\}, \varphi_{\text{effect}}^{\max} := \max_{j \in C} \{\varphi_{\text{effect}}^j\} \tag{15}$$

where $C = 2^N - 1$. (In this paper, we adopted a ‘cut one’ approximation that only evaluates $2N$ bipartitions severing the edges from a single node to the rest of the network).

The φ value of the concept as a whole or the maximally integrated cause-effect information is the minimum of the maximally integrated cause information φ_{cause} and maximally integrated effect information φ_{effect} .

$$\varphi_{\text{cause-effect}}^{\max} := \min\{\varphi_{\text{cause}}^{\max}, \varphi_{\text{effect}}^{\max}\} \tag{16}$$

If the mechanism’s maximally-irreducible cause has $\varphi_{\text{cause}} > 0$ and its maximally-irreducible effect has $\varphi_{\text{effect}} > 0$, (equivalently $\varphi_{\text{cause-effect}}^{\max} > 0$), then the mechanism is said to specify a *concept*.

We then compute the *cause-effect structure*, which is the set of all concepts specified by the subsystem characterising all of the causal constraints intrinsic to the physical system, by simply iterating the computation of the concepts over all mechanisms $M \in \mathcal{P}(S)$, where $\mathcal{P}(S)$ is the power set of subsystem nodes.

Integrated conceptual information Φ (also known as *big phi*), which is a measure of the system’s strong/integration irreducibility, is assessed by partitioning the set of elements into subsets with unidirectional cuts. The unidirectional bipartitions $P_{\rightarrow} = \{S^{(1)}, S^{(2)}\}$ of physical system S are performed by partitioning the subsystem into two parts $S^{(1)}$ and $S^{(2)}$ and cutting the

edges from one part $S^{(1)}$ to another $S^{(2)}$ (the connections are substituted with noise). We then calculate the cause–effect structure of the partitioned system $C(S^{p-})$ and compare it to $C(S)$ to evaluate the difference made by the partition. MIP, which is a search over all possible directed partitions, is then performed to identify the one that makes the least difference to the cause–effect structure. Integrated conceptual information Φ measures the irreducibility of a cause–effect structure by quantifying the difference the MIP makes to the concepts and their φ values of the system.

$$\Phi = \min_{P_{\rightarrow}} D(C(S), C(S^{p-})) \quad (17)$$

The difference D between two cause–effect structures is evaluated by an extended version of the earth mover’s distance, which is the cost of transforming one cause–effect structure $C(S)$ into another $C(S^{p-})$ in concept space.

Supporting information

S1 Fig. Heat map of $\langle\Phi\rangle$ for the rest of the parameter settings in Fig 2, where $\Delta t = 0.05$ s.
(PDF)

S2 Fig. Heat map of $\langle\Sigma\varphi\rangle$ in Fig 2 for the other parameter settings.
(PDF)

S3 Fig. Heat map of the mean $\langle\Phi\rangle$ for several parameter settings in Fig 2, where $\Delta t = 0.10$ s.
(PDF)

S4 Fig. Heat map of the mean $\langle\Phi\rangle$ of main complex and no self-loop condition for several parameter settings in Fig 2, where $\Delta t = 0.05$ s.
(PDF)

S5 Fig. Snapshot of the trajectory at $t = 100$, $t = 225$, and $t = 763$ when the IIT leadership emerges in Fig 6. The head of the group corresponds to the IIT leader.
(PDF)

S6 Fig. Match rate of the IIT leader and the MIP cut with self-loops.
(JPG)

S7 Fig. Mean $\sigma^2(\Phi)$ in Fig 2.
(PDF)

S8 Fig. Comparison of trajectories of real fish and results of the Boids model for $T = 20,000$ time steps.
(PDF)

S9 Fig. Heatmaps of the mean MI and mean TE for Boids model results with $C = 1.0$.
(PDF)

S10 Fig. Heatmaps of $\langle\Phi\rangle$ for Boids model results with different coupling strengths.
(PDF)

S11 Fig. Comparison of the averaged $\langle\Phi\rangle$. The values of Φ were averaged over the distance–visual field parameter space and again averaged for 10 simulations and all real data. Error bars indicate the standard deviations.
(PDF)

S1 Table. Transition probability matrix (TPM) used in Fig 6.
(XLSX)

Author Contributions

Conceptualization: Takayuki Niizato, Kotaro Sakamoto, Yoh-ichi Mototake.

Data curation: Takayuki Niizato, Kotaro Sakamoto, Hisashi Murakami, Takenori Tomaru, Tomotaro Hoshika, Toshiki Fukushima.

Formal analysis: Takayuki Niizato, Kotaro Sakamoto, Yoh-ichi Mototake, Hisashi Murakami.

Funding acquisition: Takayuki Niizato.

Investigation: Takayuki Niizato, Kotaro Sakamoto.

Methodology: Takayuki Niizato, Kotaro Sakamoto, Yoh-ichi Mototake.

Project administration: Takayuki Niizato, Kotaro Sakamoto.

Resources: Takayuki Niizato, Kotaro Sakamoto, Takenori Tomaru, Toshiki Fukushima.

Software: Takayuki Niizato, Kotaro Sakamoto.

Supervision: Takayuki Niizato.

Validation: Takayuki Niizato, Kotaro Sakamoto, Yoh-ichi Mototake, Hisashi Murakami.

Visualization: Takayuki Niizato, Kotaro Sakamoto.

Writing – original draft: Takayuki Niizato, Kotaro Sakamoto.

Writing – review & editing: Takayuki Niizato, Kotaro Sakamoto, Yoh-ichi Mototake, Hisashi Murakami, Takenori Tomaru.

References

1. Buhl J, Sumpter D, Couzin I, Hale J, Despland E, Miller E, et al. From disorder to order in marching locusts. *Science*. 2006; 312(5778):1402–1406. <https://doi.org/10.1126/science.1125142> PMID: 16741126
2. Yates C, Erban R, Escudero C, Couzin I, Buhl J. Inherent noise can facilitate coherence in collective swarm motion. *Proc Natl Acad Sci U S A*. 2009; 106(14):5464–5469. <https://doi.org/10.1073/pnas.0811195106> PMID: 19336580
3. Bazazi S, Bartumeus F, Hale JJ, Couzin ID. Intermittent motion in desert locusts: behavioural complexity in simple environments. *PLoS Comput Biol*. 2012. <https://doi.org/10.1371/journal.pcbi.1002498>
4. Attanasi A, Cavagna A, Castello D, Giardina I, Melillo S, Parisi L, et al. Finite-size scaling as a way to probe near-criticality in natural swarms. *Phys Rev Lett*. 2014; 113:238102. <https://doi.org/10.1103/PhysRevLett.113.238102>
5. Attanasi A, Cavagna A, Del Castello L, Giardina I, Melillo S, Parisi L, et al. Collective behaviour without collective order in wild swarms of midges. *PLoS Comput Biol*. 2014; 10(7). <https://doi.org/10.1371/journal.pcbi.1003697> PMID: 25057853
6. Murakami H, Tomaru T, Nishiyama Y, Moriyama T, Niizato T, Gunji YP. Emergent runaway into an avoidance area in a swarm of soldier crabs. *PLoS ONE*. 2014. <https://doi.org/10.1371/journal.pone.0097870>
7. Ioannou CC, Guttal V, Couzin ID. Predatory fish select for coordinated collective motion in virtual prey. *Science*. 2012. <https://doi.org/10.1126/science.1218919> PMID: 22903520
8. Strandburg-Peshkin A, Twomey CR, Bode NWF, Kao AB, Katz Y, Ioannou CC, et al. Visual sensory networks and effective information transfer in animal groups. *Curr Biol*. 2013. <https://doi.org/10.1016/j.cub.2013.07.059> PMID: 24028946
9. Berdahl A, Torney CJ, Ioannou CC, Faria JJ, Couzin ID. Emergent sensing of complex environments by mobile animal groups. *Science*. 2013. <https://doi.org/10.1126/science.1225883> PMID: 23372013

10. Murakami H, Niizato T, Tomaru T, Nishiyama Y, Gunji YP. Inherent noise appears as a Lévy walk in fish schools. *Sci Rep*. 2015. <https://doi.org/10.1038/srep10605>
11. Niizato T, Murakami H, Sangu K, Tomaru T, Sonoda K, Nishiyama Y, et al. Local perspectives of *Plecoglossus altivelis* determine searching strategy. In: *AIP Conf Proc*. 2017.
12. Ballerini M, Cabibbo N, Candelier R, Cavagna A, Cisbani E, Giardina I, et al. Empirical investigation of starling flocks: a benchmark study in collective animal behaviour. *Anim Behav*. 2008. <https://doi.org/10.1016/j.anbehav.2008.02.004>
13. Cavagna A, Cimarelli A, Giardina I, Parisi G, Santagati R, Stefanini F, et al. Scale-free correlations in starling flocks. *Proc Natl Acad Sci*. 2010. <https://doi.org/10.1073/pnas.1005766107> PMID: 20547832
14. Cavagna A, Queirós SMD, Giardina I, Stefanini F, Viale M. Diffusion of individual birds in starling flocks. *Proc Biol Sci*. 2013.
15. Bialek W, Cavagna A, Giardina I, Mora T, Pohl O, Silvestri E, et al. Social interactions dominate speed control in poising natural flocks near criticality. *Proc Natl Acad Sci*. 2014. <https://doi.org/10.1073/pnas.1324045111>
16. Attanasi A, Cavagna A, Del Castello L, Giardina I, Jelic A, Melillo S, et al. Emergence of collective changes in travel direction of starling flocks from individual birds' fluctuations. *J R Soc Interface*. 2015. <https://doi.org/10.1098/rsif.2015.0319> PMID: 26236825
17. Mora T, Walczak AM, Del Castello L, Ginelli F, Melillo S, Parisi L, et al. Local equilibrium in bird flocks. *Nat Phys*. 2016. <https://doi.org/10.1038/nphys3846> PMID: 27917230
18. Rosenblatt F. The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain. *Psychol. Rev*. 1958. <https://doi.org/10.1037/h0042519> PMID: 13602029
19. Hopfield J. Neural networks and physical systems with emergent collective computational abilities. *PNAS*. 1982. <https://doi.org/10.1073/pnas.79.8.2554> PMID: 6953413
20. Hesse J. and Gross T. Self-organized criticality as a fundamental property of neural systems. *Front. Syst. Neurosci*. 2014. <https://doi.org/10.3389/fnsys.2014.00166> PMID: 25294989
21. Lynn CW., Bassett DS. The physics of brain network structure, function and control. *Nat. Phys*. 2019.
22. Langton C. Computation at the edge of chaos: Phase transitions and emergent computation. *Physica D*. 1990. [https://doi.org/10.1016/0167-2789\(90\)90064-V](https://doi.org/10.1016/0167-2789(90)90064-V)
23. Pennington J., Schoenholz S., Ganguli S. The emergence of spectral universality in deep networks. *arXiv:1802.09979 AISTATS*. 2018.
24. Xiao L. and Bahri Y. and Sohl-Dickstein J. and Schoenholz S. and Pennington J. Dynamical Isometry and a Mean Field Theory of CNNs: How to Train 10,000-Layer Vanilla Convolutional Neural Networks. *arXiv:1806.05393 ICML*. 2018.
25. Franks NR, Dornhaus A, Fitzsimmons JP, Stevens M. Speed versus accuracy in collective decision making. *Proc Biol Sci*. 2003. <https://doi.org/10.1098/rspb.2003.2527>
26. Dyer JRG, Johansson A, Helbing D, Couzin ID, Krause J. Leadership, consensus decision making and collective behaviour in humans. *Philos Trans R Soc Lond B Biol Sci*. 2009. <https://doi.org/10.1098/rstb.2008.0233> PMID: 19073481
27. Bose T, Reina A, Marshall JA. Collective decision-making. *Curr Opin Behav Sci*. 2017. <https://doi.org/10.1016/j.cobeha.2017.03.004>
28. Couzin ID, Ioannou CC, Demirel G, Gross T, Torney CJ, Hartnett A, et al. Uninformed individuals promote democratic consensus in animal groups. *Science*. 2011. <https://doi.org/10.1126/science.1210280> PMID: 22174256
29. Pinkoviezky I, Couzin ID, Gov NS. Collective conflict resolution in groups on the move. *Phys Rev E*. 2018. <https://doi.org/10.1103/PhysRevE.97.032304> PMID: 29776127
30. Couzin I. Collective minds. *Nature*. 2007. <https://doi.org/10.1038/445715a> PMID: 17301775
31. Couzin ID. Collective cognition in animal groups. *Trends Cogn Sci*. 2009; 13(1):36–43. <https://doi.org/10.1016/j.tics.2008.10.002> PMID: 19058992
32. Sumpter D. *Collective animal behavior*. Princeton University Press; 2010.
33. Bak P, Tang C, Wiesenfeld K. Self-organized criticality. *Phys Rev A*. 1988. <https://doi.org/10.1103/PhysRevA.38.364>
34. Tetzlaff C, Okujeni S, Egert U, Wörgötter F, Butz M. Self-organized criticality in developing neuronal networks. *PLoS Comput Biol*. 2010. <https://doi.org/10.1371/journal.pcbi.1001013> PMID: 21152008
35. Niizato T, Gunji YP. Fluctuation-driven flocking movement in three dimensions and scale-free correlation. *PLoS ONE*. 2012. <https://doi.org/10.1371/journal.pone.0035615> PMID: 22662109
36. Gunji YP. Self-organized criticality in asynchronously tuned elementary cellular automata. *Complex Syst*. 2014. <https://doi.org/10.25088/ComplexSystems.23.1.55>

37. Gunji YP, Sakiyama T, Murakami H. Punctuated equilibrium based on a locally ambiguous niche. *BioSystems*. 2014. <https://doi.org/10.1016/j.biosystems.2014.05.005>
38. Niizato T, Murakami H. Entangled time in flocking: multi-time-scale interaction reveals emergence of inherent noise. *PLoS ONE*. 2018; 13(4):1–21. <https://doi.org/10.1371/journal.pone.0195988>
39. Herbert-Read JE, Krause S, Morrell LJ, Schaerf TM, Krause J, Ward AJW. The role of individuality in collective group movement. *Proc Biol Sci*. 2013. <https://doi.org/10.1098/rspb.2012.2564> PMID: 23222452
40. Jolles JW, Boogert NJ, Sridhar VH, Couzin ID, Manica A. Consistent individual differences drive collective behavior and group functioning of schooling fish. *Curr Biol*. 2017. <https://doi.org/10.1016/j.cub.2017.08.004> PMID: 28889975
41. Katz Y, Tunstrom K, Ioannou CC, Huepe C, Couzin ID. Inferring the structure and dynamics of interactions in schooling fish. *Proc Nat Acad Sci*. 2011. <https://doi.org/10.1073/pnas.1107583108>
42. Gautrais J, Ginelli F, Fournier R, Blanco S, Soria M, Chaté H, et al. Deciphering interactions in moving animal groups. *PLoS Comput Biol*. 2012. <https://doi.org/10.1371/journal.pcbi.1002678> PMID: 23028277
43. Staniek M, Lehnertz K. Symbolic transfer entropy. *Phys Rev Lett*. 2008. <https://doi.org/10.1103/PhysRevLett.100.158101> PMID: 18518155
44. Butail S, Mwaffo V, Porfiri M. Model-free information-theoretic approach to infer leadership in pairs of zebrafish. *Phys Rev E*. 2016. <https://doi.org/10.1103/PhysRevE.93.042411> PMID: 27176333
45. Crosato E, Jiang L, Lecheval V, Lizier JT, Wang XR, Tichit P, et al. Informative and misinformative interactions in a school of fish; Swarm Intel. 2018; 12(4):283–305. <https://doi.org/10.1007/s11721-018-0157-x>
46. Lizier JT, Prokopenko M, Zomaya AY. Local information transfer as a spatiotemporal filter for complex systems. *Phys Rev E*. 2008. <https://doi.org/10.1103/PhysRevE.77.026110>
47. Lizier JT, Prokopenko M, Zomaya AY. Local measures of information storage in complex distributed computation. *Inf Sci*. 2012. <https://doi.org/10.1016/j.ins.2012.04.016>
48. Sun J, Bollt EM. Causation entropy identifies indirect influences, dominance of neighbors and anticipatory couplings. *Physica D*. 2014. <https://doi.org/10.1016/j.physd.2013.07.001>
49. James RG, Barnett N, Crutchfield JP. Information flows? A critique of transfer entropies. *Phys Rev Lett*. 2016. <https://doi.org/10.1103/PhysRevLett.116.238701>
50. Tomaru T, Murakami H, Niizato T, Nishiyama Y, Sonoda K, Moriyama T, et al. Information transfer in a swarm of soldier crabs. *Artif Life Robot*. 2016. <https://doi.org/10.1007/s10015-016-0272-y>
51. Erten E, Lizier JT, Piraveenan M, Prokopenko M. Criticality and information dynamics in epidemiological models. *Entropy*. 2017. <https://doi.org/10.3390/e19050194>
52. Wang XR, Miller JM, Lizier JT, Prokopenko M, Rossi LF. Quantifying and tracing information cascades in swarms. *PLoS ONE*. 2012.
53. Wang XR, Miller JM, Lizier JT, Prokopenko M, Rossi LF. Measuring information storage and transfer in swarms. In: *Proc Eleventh Eur Conf Synth Simul Living Syst (ECAL 2011)*. 2011.
54. Albantakis L, Tononi G. The intrinsic cause–effect power of discrete dynamical systems—from elementary cellular automata to adapting animats. *Entropy*. 2015. <https://doi.org/10.3390/e17085472>
55. Christensen WD B M. The process dynamics of normative function. *Monist*. 2002; 85(1):3–28. <https://doi.org/10.5840/monist20028516>
56. Tononi G. Information measures for conscious experience *Arch Ital Biol*. 2001; 4(31): 367–371
57. Tononi G., Sporns O. Measuring information integration *BMC Neuroscience*. 2003; 129(4)
58. Balduzzi D, Tononi G. Qualia: the geometry of integrated information. *PLoS Comput Biol*. 2009. <https://doi.org/10.1371/journal.pcbi.1000462> PMID: 19680424
59. Tononi G. An integrated information theory of consciousness. In: *Encyclopedia of Consciousness*; 2010.
60. Barrett AB, Seth AK. Practical measures of integrated information for time-series data. *PLoS Comput Biol*. 2011. <https://doi.org/10.1371/journal.pcbi.1001052> PMID: 21283779
61. Oizumi M, Albantakis L, Tononi G. From the phenomenology to the mechanisms of consciousness: integrated information theory 3.0. *PLoS Computat Biol*. 2014. <https://doi.org/10.1371/journal.pcbi.1003588>
62. Pearl J. Causality: models, reasoning, and inference. *IIE Transactions*. 2002; 34(6):583–589.
63. Ay N, Polani D. Information flows in causal networks. *Adv Complex Syst*. 2008; 11(01):17–41. <https://doi.org/10.1142/S0219525908001465>

64. Massimini M, Ferrarelli F, Huber R, Esser SK, Singh H, Tononi G. Neuroscience: Breakdown of cortical effective connectivity during sleep. *Science*. 2005.
65. Alkire MT, Hudetz AG, Tononi G. Consciousness and anesthesia. *Science (New York, NY)*. 2008. <https://doi.org/10.1126/science.1149213>
66. Gosseries O, Di H, Laureys S, Boly M. Measuring Consciousness in Severely Damaged Brains. *Annual Review of Neuroscience*. 2014. <https://doi.org/10.1146/annurev-neuro-062012-170339> PMID: 25002279
67. Khajehabdollahi S OASA Abeyasinghe P. The emergence of integrated information, complexity, and consciousness at criticality. *BioRxiv*:1606.08313 [Preprint].2019.
68. Mediano PAM SM Farah JC. Integrated information and metastability in systems of coupled oscillators. *arXiv*:1606.08313 [Preprint]. 2016.
69. Mori H, Oizumi M. Information integration in a globally coupled chaotic system. 2018.
70. Oizumi M, Tsuchiya N, Amari S. A unified framework for information integration based on information geometry. 2015; 113(51). *PNAS* <https://doi.org/10.1073/pnas.1603583113>
71. Oizumi M, Amari SI, Yanagawa T, Fujii N, Tsuchiya N. Measuring integrated information from the decoding perspective. *PLoS Computat Biol*. 2016. <https://doi.org/10.1371/journal.pcbi.1004654>
72. Mayner W, Marshall W, Albantakis L, Findlay G, Marchman R, Tononi G. PyPhi: a toolbox for integrated information theory. *PLoS Comput Biol*. 2018. <https://doi.org/10.1371/journal.pcbi.1006343> PMID: 30048445
73. Mediano P, Rosas F, Carhart-Harris R, Seth A, Barrett A Beyond integrated information: A taxonomy of information dynamics phenomena *arXiv*:1909.02297 [Preprint]. 2019.
74. Mediano P, Seth A, Barrett A. Measuring integrated information: comparison of candidate measures in theory and simulation. 2019; *Entropy*. 2019.
75. Ito S. Stochastic thermodynamic interpretation of information geometry. *Phys Rev Lett*. 2018. <https://doi.org/10.1103/PhysRevLett.121.030605>
76. Ito S. Unified framework for the second law of thermodynamics and information thermodynamics based on information geometry. *arXiv*:1810.09545 [Preprint]. 2018.
77. Farnsworth K. How organisms gained causal independence and how it might be quantified. *Biology*. 2018. <https://doi.org/10.3390/biology7030038> PMID: 29966241
78. Hooker C. Philosophy of complex systems (Handbook of the philosophy of science, vol. 10). North Holland; 2011.
79. Bertalanffy L. General system theory: foundations, development, applications (Revised Edition). George Braziller; 1969.
80. Albantakis L, Marshall W, Hoel E, Tononi G. What caused what? A quantitative account of actual causation using dynamical causal networks. *Entropy*. 2019; 21(5):1099–4300. <https://doi.org/10.3390/e21050459>
81. Edlund JA, Chaumont N, Hintze A, Koch C, Tononi G, Adami C. Integrated information increases with fitness in the evolution of animats. *PLoS Comput Biol*. 2011. <https://doi.org/10.1371/journal.pcbi.1002236> PMID: 22028639
82. Kanakov O, Gordleeva S, Ermolaeva A, Jalan S, Zaikin A. Astrocyte-induced positive integrated information in neuron-astrocyte ensembles. *Phys Rev E*. 2019; 99:012418. <https://doi.org/10.1103/PhysRevE.99.012418> PMID: 30780273
83. Marshall W, Kim H, Walker SI, Tononi G, Albantakis L. How causal analysis can reveal autonomy in models of biological systems. *Philos Trans A Math Phys Eng Sci* 2017. <https://doi.org/10.1098/rsta.2016.0358>
84. Engel David M TW. Integrated information as a metric for group interaction. *Artif Life Robot*. 2018.
85. Aguilera M, Alquézar C, Bedia MG. Agency and integrated information in a minimal sensorimotor model; 2018.
86. Aguilera M, Paolo ED. Integrated information in the thermodynamic limit. 2019; *Neural Netw*. 2019.
87. Esteban FJ LJPJSTF Galadí JA. Informational structures: a dynamical system approach for integrated information. *PLOS Comput Biol*. 2018; 14: e1006154. <https://doi.org/10.1371/journal.pcbi.1006154> PMID: 30212467
88. Couzin ID, Krause J, James R, Ruxton GD, Franks NR. Collective memory and spatial sorting in animal groups. *J Theor Biol*. 2002. <https://doi.org/10.1006/jtbi.2002.3065> PMID: 12297066
89. Vicsek T, Zafeiris A. Collective motion *Phys Rep*. 2012; 517(3–4):71–140.
90. Niizato T, Sakamoto K, Mototake Y., Murakami H, Tomaru T, Hoshika T, et al. New Classification of Collective Animal Behaviour as an Autonomous System. *Preprints*:201902.0269 [Preprint].2019.

91. Nagy M, BDVT Ákos Z. Hierarchical group dynamics in pigeon flocks. *Nature*. 2010; 464: 890–3. <https://doi.org/10.1038/nature08891>
92. Ozogány K, Vicsek T. Modeling the Emergence of Modular Leadership Hierarchy During the Collective Motion of Herds Made of Harems. *Journal of Statistical Physics*. 2015; 158:628–646.
93. Jia Y, VT. Modelling Hierarchical Flocking. arXiv:1810.09545 [Preprint]. 2018.
94. Ishiwata R., Kinukawa R., and Sugiyama Y. Analysis of dynamically stable patterns in a maze-like corridor using the Wasserstein metric *Scientific Reports*. 2018; 8:6367
95. CW R. Flocks, herds, and schools: a distributed behavioral model. *SIGGRAPH Comput Graph*. 1987; 21(4):25–34. <https://doi.org/10.1145/37402.37406>
96. Amari S., Tsuchiya N., Oizumi M. Geometry of Information Integration. In: Ay N., Gibilisco P., Matús F. (eds) *Information Geometry and Its Applications*. IGAIA IV 2016. Springer Proceedings in Mathematics & Statistics, vol 252. Springer, Cham
97. Berlow EL. Strong effects of weak interactions in ecological communities. *Nature*. 1999; 398(25):330–334. <https://doi.org/10.1038/18672>
98. Kadoya T, McCann KS. Weak Interactions and Instability Cascades. *Scientific Reports*. 2015. <https://doi.org/10.1038/srep12652> PMID: 26219561
99. Landi P, Minoarivelo HO, Brännström Å, Hui C, Dieckmann U. Complexity and stability of ecological networks: a review of the theory; 2018.
100. Olfati-Saber R. Flocking for multi-agent dynamic systems: algorithms and theory. *IEEE Transact Automat Contr*. 2006. <https://doi.org/10.1109/TAC.2005.864190>
101. Olfati-Saber R, Fax JA, Murray RM. Consensus and cooperation in networked multi-agent systems. *Proc IEEE*. 2007. <https://doi.org/10.1109/JPROC.2006.887293>