



HAL
open science

Identifying interaction neighbours in animal groups

Hadrien-Matthieu Gascuel, Fernando Peruani, Richard Bon

► **To cite this version:**

Hadrien-Matthieu Gascuel, Fernando Peruani, Richard Bon. Identifying interaction neighbours in animal groups. *Animal Behaviour*, 2021, 174, pp.97-104. 10.1016/j.anbehav.2021.01.019 . hal-03385453

HAL Id: hal-03385453

<https://hal.science/hal-03385453>

Submitted on 10 Mar 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

Identifying interaction neighbours in animal groups

Hadrien-Matthieu Gascuel^{1, 2}, Fernando Peruani^{2, 3, *}, Richard Bon^{1, **}

¹Centre de Recherches sur la Cognition Animale (CRCA), Toulouse, France

²Université Côte d'Azur, Laboratoire J. A. Dieudonné, Nice, France.

³Laboratoire de Physique Théorique et Modélisation, Cergy Paris Université, France

* Correspondence: Laboratoire de Physique Théorique et Modélisation, CNRS UMR 8089, CY Cergy Paris Université, F-95302 Cergy-Pontoise Cedex, France

** Correspondence: Centre de Recherches sur la Cognition Animale (CRCA), Centre de Biologie Intégrative (CBI), Université de Toulouse; CNRS, UPS, France

E-mail addresses: fernando.peruani@cyu.fr (F. Peruani), richard.bon@univ-tlse3.fr (R. Bon)

Word count: 7469

Abstract

Identifying interaction partners – a particularly challenging task in the absence of an explicit communication system (e.g. vocalization) – is key to understand how information is distributed and processed within animal groups. Moreover, the interaction network (IN) regulates complex collective behaviours such as collective motion and predator detection. Despite the relevance of the IN, there are only few experiments specially designed to decipher it. Moreover, the mechanism by which an

individual selects its interaction neighbours remains largely unknown. Here, we investigate the interaction-neighbour selection process, by studying the behavioural response of naive group members confronted to the controlled departure of a trained individual (TI) in herds of merino sheep. By identifying the first individual that follows the TI, we infer the probability for a group member to choose the TI as interaction neighbour, assuming that this probability is modulated by either the relative distance, the relative distance-rank, or a combination of both. The underlying IN is constructed from this probability. The obtained results provide unique, reliable information on the functional form of this probability, and thus on the IN and on how individuals interact within groups. The suggested method to unveil the selection process can be easily implemented in other animal systems and contexts.

Keywords: controlled departure, follower, interaction setting, Merino sheep, neighbourhood, probabilistic model

Large groups of animals such as fish shoals, bird flocks, and sheep herds are able to produce captivating collective motion patterns, preserving cohesion, while performing complex manoeuvres or escaping predators (Ginelli et al., 2015; Parrish & Edelstein-Keshet, 1999; Sumpter, 2006). The spatio-temporal coordination of large moving groups is assumed to be the result of local interactions (Camazine et al. 2001). Interactions or mutual stimulations are basic ingredients of social living systems spanning from cells to societies. Social interactions are involved in aggregation (Grassé, 1946; Lesne & Jeanson, 2015), known to influence phenotypes (Grassé 1968), and play a central role in the organization of sexual displays (Chauvin, 1982; Tinbergen, 1963; Wyatt, 2003), choice of reproductive sites (Doligez et al., 2002), search of food sources (Giraldeau & Caraco, 2000, Lihoreau & Rivault, 2011; Valone & Templeton, 2002), shelter selection (Amé et al., 2006), and behaviour synchrony (Bloch et al., 2013). Besides, animals' propensity to lead varies according to their dominance status (Sueur et al., 2012), energetic requirements (Conradt et al., 2009; Fischhoff et al., 2007), while the propensity to follow may depend on friendship (Sueur et al., 2009, 2010), or body size (Reebs, 2001). Leader and follower behaviour can also result from motion characteristics such as speed and straightness of displacement (Couzin et al., 2002a; Jolles et al., 2020; Tang et al., 2020). Finally, environmental heterogeneity is known to modulate the density of animals and influences the global structure and motion of groups (Casellas et al., 2008; Chepizhko et al., 2013).

Studying collective phenomena through the scope of self-organization implies that complex, global patterns emerge, in large group sizes that exceed the perceptual and/or cognitive capacity of individuals, from local interactions. Thus, the existence of an “interaction neighbourhood” for each group member is assumed. For a given focal individual (FI), the group members it can receive information from and to which it can respond to (e.g. approach, avoid, follow) are the interaction neighbours. Identifying the interaction neighbours of a FI is fundamental to understand how information is processed and distributed through the group (Buhl et al., 2006; Lima, 1995; Rosenthal et

al., 2015). The importance of identifying the interaction neighbours outstrips the study of animal behaviour as it is of interest in other scientific fields as in the study of social networks, social sciences, management and economics (Kim, 2014).

There is evidence that in animal groups the interaction neighbours of a FI – in the absence of a strong hierarchical social structure – are those group members located at close proximity of the FI (Ballerini et al., 2008; Herbert-Read et al., 2011; Reuter & Breckling, 1994; Viscido et al., 2002, 2005). The degree of influence of an interaction neighbour is often assumed to be a decreasing function of the relative distance between the two individuals (Heras et al., 2019; Herbert-Read et al., 2011; Katz et al., 2011; Torney et al., 2018). There exist several options on how the degree of influence varies with the distance. One option is to assume that given two individuals, the influence depends on the Euclidean distance between them. Another option is to assume that the closest individual is the most influential, the second closest individual, the second most influential, and so on, in such a way that the influence is based on a distance rank rather than directly on the Euclidean distance. If a distance rank is used, the influence of the closest individual results the same if this individual is located at a distance of e.g. 5m or 10m, while for a degree of influence based on the Euclidean distance, these two situations lead to different values.

Several mathematical interaction neighbour models (INMs), based on both Euclidean distance and distance rank, have been proposed (Aoki, 1982; Ballerini et al., 2008; Herbert-Read et al., 2011; Kunz & Hemelrijk, 2012; Reuter & Breckling, 1994; Rosenthal et al., 2015; Viscido et al., 2005). The simplest Euclidean INM, extensively used in agent-based, social-force models (Couzin et al., 2002b; Vicsek et al., 1995), assumes the existence of a threshold R , such that all individuals located at a distance smaller or equal to R from the FI are interaction neighbours and exhibit the same degree of influence, while those at a distance larger than R are not interaction neighbours. More realistic extensions of this simple model assume that perception is limited by a blind zone, which requires the use of an interaction cone (Aoki, 1982, Barberis & Peruani, 2016). Finally, using also social-force

models, it was explored in Reuter & Breckling (1994) an influence, between two individuals, that goes as the inverse of their relative distance, while Viscido et al. (2002, 2005) tested a series of mathematical functions to assess the effect of the distance-dependent influence. The idea that an INM has to be necessarily based on the relative Euclidean distance between individuals was challenged by Ballerini et al. (2008). Based on empirical evidence, Ballerini et al. (2008) showed that starlings flying in a flock interact with a (roughly) fixed number of nearest (in the Euclidean sense) neighbours and independently of the local density. As pointed out by Ballerini et al., this is in sharp contrast with what is expected for metric interactions (Kunz & Hemelrijk, 2012) and refer to this set of neighbours, i.e. this fixed number of nearest neighbours, as “topological” neighbours. Topological neighbours can be identified by performing a Voronoi Tessellation or by performing a distance ranking for each individual. The use of topological neighbours ensures flock cohesion, even if the flock experiences large density fluctuations, and facilitates information spreading, which takes the form, for the analyzed birds, of an undamped, linearly propagating wave (Attanasi et al., 2014).

Here, we investigate the process by which individuals select interaction neighbours. We characterize this process by the probability $W(A, B)$ that B chooses A as interaction neighbour. Specifically, we look at a generic activation process, where initially all individuals are inactive (sharing the same behavioural state), except for one individual (the initiator) that is activated (undergoes a behavioural change). Our goal is to determine $W(A, B)$ from experimental data by studying the relative position of the first individual, after the initiator, to get activated, and for whom, we know with certainty, the initiator is an interaction neighbour (Fig. 1a and 1b). By focusing on the first activation, we avoid the ambiguity that results from having multiple activators as well as possible nonlinear effects that may lead, for instance, to an acceleration of the process as the number of active individuals increases over time (Toulet et al. 2015). We assume that $W(A, B)$ can depend on the Euclidean distance d_{MAB} , distance rank d_{TAB} , or a combination of both. Note that $d_{MAB} = d_{MBA}$, while in general $d_{TAB} \neq d_{TBA}$, since d_{TAB} is a rank and not a distance, as illustrated in Fig. 1c, 1d, 1e and 1f. As a direct consequence of

this, if $W(A, B)$ is based on Euclidean distance, then $W(A, B) = W(B, A)$, while when based on a distance rank we have to consider that in general $W(A, B) \neq W(B, A)$. Thus, the resulting interaction network tends to exhibit undirected links when using Euclidean distance, and directed ones for distance rank (Fig. 1g and 1h). This strongly affects how information, here an activation, propagates over the system. To illustrate these ideas, we investigate $W(A, B)$ in sheep herds. More specifically, we study the behavioural response of naive group members, initially static, confronted with the experimentally controlled departure of a trained individual (the initiator). In this context, an activation occurs when a naive individual starts following the trained individual. By identifying the first individual that follows the initiator, we determine $W(A, B)$. The procedure outlined here might prove of use in other animal systems as well as in contexts different from a collective departure.

MATERIAL AND METHODS

Study area and sheep herd

Experiments were conducted at the Domaine du Merle, an experimental farm in the south of France. This field station covers 400 ha within a native steppe in the old delta of the Durance river. This area is particularly flat. A system of irrigation permits to obtain homogeneous pastures. The herd of merino sheep raised in the station graze these pastures in autumn and winter. Later they forage in the native steppe and hills around the field station before being transported in the Alps for the summer period. To conduct our experiments, we randomly choose from the available herd of 1400 ewes, 30 sheep at random to be trained and 200 sheep used as naive individuals during the experiments. All animals were 3 to 6 years old. The aim was to provoke departures of one individual in groups of 32 grazing sheep. This was possible by previously training some sheep to move toward a coloured panel raised at distance when stimulated by vibrating collar.

Training of initiators and habituation of naive sheep

Training of future initiators was performed using 6 stable groups of 5 individuals each, following the protocol described in (Pillot et al., 2011, 2010; Toulet et al., 2015). Sheep were trained to associate a vibration delivered by a neck collar and a food reward (handful of corn) delivered by a panel raised at the same time at the periphery of one 50 × 50 m arena. After the training phase we selected the 3 sheep that exhibited the best learning scores (100 % of departures towards the panel rising following a stimulation) to proceed with the experiments. Meanwhile, the set of naive sheep (N = 200) to be used was habituated to panel rising (without corn delivery) at the periphery of experimental arenas, at one-min interval during two sessions of 90 min. At the end of this habituation session, none of the 200 naive raised the head when the panel was lifted up.

Trials of collective movements and control trials

The trials of collective movements consisted in introducing 32 sheep groups (1 trained sheep and 31 naives) within one of the two arenas delimited with sheep fences. All experiments were carried out from 10 AM to 6 PM and all sheep were penned up in the same sheepfold and fed hay in the morning and the evening. Each trial lasted 30 min. Once introduced within one arena, sheep were freely grazing during 20 min. Then the trained sheep was stimulated (2 s) through a vibrating neck collar (activated remotely) while at the same time one of two coloured panels located at the external periphery of the arena was raised. The stimulation was delivered when all sheep were grazing head down. The trained individual did not show preferences in term of positions as it has been observed, relatively to panel raised, either in front, in the middle or in the back of the group. Past 10 min (end of test), a new panel was raised (one of the two farthest from the tower) to reinforce the conditioning of trained ewes and

avoid restricting their space use to the vicinity of panels closest to the tower. The group was led back to the sheepfold shortly afterward. We performed 30 experimental trials. The naive sheep that were not tested during one experimental day were introduced in distant pasture.

We also carried out control experiments different from the habituation sessions to be sure that naive ewes did not associate the panel rise and the food reward. Thereby, 6 trials before and 6 after the experimental trials have been conducted with groups of 32 naive ewes, using the same protocol as described before. We found no movement of naive individuals during these 12 control trials.

Data Collection and analyses

A 7 m high tower was located near the two arenas. Two digital cameras (Canon EOS D50) were anchored on the top of the tower, each one focusing on one arena. Fifteen minutes after the introduction of the 32-sheep groups, each containing 1 trained and 31 naive individuals, the digital camera was turned on, taking a picture of the entire arena every second and turned off five minutes after the stimulation of the trained sheep. For each trial, about 600 pictures were collected. Using a custom software, we track on each picture the position of animals by dragging a vector on their back, and identify the behaviour of each individual, i.e. grazing, standing head-up, moving and others. Due to the oblique viewing angle, we use an algorithm detailed in Ginelli et al. (2015) to compute the real positions. We defined the departure of the trained sheep (initiator), as an uninterrupted walk towards the raised panel immediately following the experimental stimulation (neck collar vibration). The first follower was identified as the first naive sheep that moved after the trained sheep departure, without stopping until joining the trained ewe near the panel. The identification of the first followers was possible because the followings were successive, even if followers departed almost simultaneously at many occasions. Twelve replications were discarded, one because the initiator did not depart, five

because the behaviour of the initiator was atypical and not comparable to other trials and six because no naive sheep followed the initiator's departure. The analyses were thus performed on 18 trials. From the 18 trials selected to perform the analyses, we extracted the configuration of the group at the moment the initiator departed from the group i.e. the position of all individuals in the arena. In each trial with collective following, we were able to identify the first follower. The location of each sheep allowed calculating the inter-distance between all sheep and so ranking all sheep by their metric distance and rank of distance to each individual in a group. These distances were calculated using the location of sheep the second before the departure of the trained sheep.

Details on the simulations

Simulations were created in order to assess which INM performed better in comparison to experimental data. Obtaining all relative distances and distance ranks from the experiments, we simulated the first activation that follows the departure of the trained individual by applying a Monte Carlo algorithm using the test function $W(A, B)$. For each (field) experiment, 500 simulations were performed to compute the probability for each naive individual to act as first follower, and the cumulative distribution of distances between the initiator and the simulated first follower. Parameter estimation was performed by maximum likelihood over the 18 experiments. The statistical relevance of each tested function $W(A, B)$ was assessed by performing the 2-sample Kolmogorov-Smirnov test (James, 2006). This statistical test compares the maximum "distance" between two cumulative distributions with a critical tolerable distance given by the expression:

$$D_{critical}(n, m, \alpha) = \frac{1}{\sqrt{n}} \sqrt{-\ln(\alpha/2) \frac{1+\frac{n}{m}}{2}},$$

where n and m are the size of each sample, and α is the significant level. Along the study, we used a standard significant level of $\alpha = 0.05$.

Ethics statement

All the animals were maintained under routine husbandry conditions at a Montpellier Supagro research station (Domaine du merle, Salon-de-Provence, France) with full approval of its director Pierre-Marie Bouquet. Animal welfare requirements were fully respected in accordance with the European Directive 2010/63/EU, with the rules of the European Convention for the Protection of Vertebrate Animals used for Experimental and Other Scientific Purposes and with the Convention of the French Comité national de réflexion éthique sur l'expérimentation animale. No special authorization from the French Ethical Committee for animal experimentation (Commission nationale de l'expérimentation animale) was required as no protected or endangered species was involved, as the experiments did not imply any invasive manipulation (the experimental protocol consists in the observation of groups and the acquired data are only pictures of the animals) and as sheep were conducted to the test arenas, as they are herded on a daily basis to the pastures. All personnel involved had technical support from the employees of the research station as required by the French Ministry of Research. The experimental protocols included short test periods (35 minutes) where sheep did not experience painful, stressful or unfamiliar situations. The experimental procedures had no detrimental effect on the sheep and at the end of the experiment all the animals reintegrated the sheep herd of the breeding research station.

RESULTS

As explained in the introduction, we characterized the interaction-neighbour selection process via the probability $W(A, B)$ that B chooses A as interaction neighbour. Our strategy is to determine $W(A, B)$ using a simplified scenario where all group members are initially in the same behavioural state, and then force a behavioural change in one individual to identify the first individual that responds to the initiator. Note that in this scenario there is no ambiguity to whom this first individual is responding to. Our experimental system consist of a group of $N - 1$ naive sheep and one trained individual (here, $N = 32$). When all individuals are in a motionless phase, the trained individual is stimulated to move to a given location in the field by activating, remotely, a vibrating collar. We assume naive sheep do not perceive the activation of the collar. We identify the first individual that responds to the activation of the trained individual, i.e. the first individual that starts moving toward the trained individual. The experiment is repeated using different sets of 31 individuals, and performed only once with each naive group; for more details see Material and Methods. To determine $W(A, B)$, we test a series of functional forms for the probability $W(A, B)$, in increasing order of complexity or number of parameters. Our goal is to determine which hypotheses are consistent with the experimental data. For that purpose, we perform a large number of in silico experiments, in which the location of each individual is taken from the experiments and using the proposed functional form of $W(A, B)$, we simulate which individual is the first to select trained individual (initiator) and imitates its behaviour (i.e. moves in the same direction). We compute from the simulated data the cumulative distribution (CDD) of the inter-distance between the first followers and the initiator. This distribution indicates the probability that the first follower is found at a distance smaller than a given distance from the initiator. Finally, we compare the experimental and simulated CDD.

Global Perception – a 0 parameter model

We start out by testing the simplest hypothesis – we call it “global perception” – by which we assume that all individuals exhibit the same probability to choose as interaction neighbour the initiator:

$$W(A, B) = \frac{1}{N-1} \quad (1)$$

This model assumes that each individual perceives all conspecifics and can select, with the same probability, any individual in the system as interaction neighbour, irrespective of its location. This model is reasonable, when dealing with small group sizes, where it can be ensured that the cognitive and perception capacity of individuals allow them to interact with any group member. In this limit, no spatial effects are observable. For group sizes that exceed the perception and/or cognitive capacity of the individuals, the global perception model is inadequate. Using $N = 32$ sheep, deviations between the global perception model and the experiments, though weak, are statistically significant. The Kolmogorov-Smirnov test indicates that this model is not consistent with the experimental data [D-statistics, $D = 0.34 > D_{\text{critical}, 0.05} = 0.33$]; as illustrated by the corresponding CDD (Fig. 2a). This implies, as suggested by the experimental data (Fig. 1a and 1b), that $W(A, B)$ depends on the relative position between individual A and B . This possibility is investigated below.

Distance and Distance Rank models – 1 parameter model

Now, we study models that depend on only R_0 one parameter. We assume that $W(A, B)$ depends on either the Euclidean distance or the distance rank between A and B . We start out by exploring interaction-neighbour selection models inspired on the most common metric (Euclidean distance) and topological (distance rank) neighbourhood models used in the literature. We express the probability, inspired on the standard metric model, as:

$$W(A, B) = \begin{cases} \frac{1}{n} & \text{if } d_{MBA} \leq R_0 \\ 0 & \text{if } d_{MBA} > R_0 \end{cases} \quad (2)$$

Where d_{MBA} is the Euclidean distance between B and A , R_0 is a constant that defines the interaction radius, and the only parameter of the model, and n is the number of individuals for which A is at a distance less than or equal to R_0 . Thus, the number n depends on the group density. Note that all individuals within the interaction radius exhibit the same probability of choosing A as interaction neighbour. It is important to note that according to this definition of $W(A, B)$ for small values of R_0 , even if all interaction neighbours are included, the resulting interaction network is not percolated, while for large values, the network is percolated. Since we consider experiments in which the departure of the trained individual leads to a collective departure, the topology of the interaction network should be such that a global activation can be ensured. This means that the interaction network permits that each individual is connected to at least one group member as illustrated in Fig. 1g. In practice, this implies that R_0 has to be $R_0 > R^*$, where R^* is the critical value of R_0 (see Fig. 2b). Provided percolation is ensured, we estimate by maximum likelihood estimation (MLE), the value of R_0 that best describes the experimental data. The obtained CDD, using this value of R_0 , is shown in Fig. 2d. A KS -test indicates this model is consistent with the experimental data [D -statistics, $D = 0.31 < D_{\text{critical},0.05} = 0.33$]. Now, we focus on a probability inspired on the standard topological model as:

$$W(A, B) = \begin{cases} \frac{1}{k} & \text{if } A \text{ within } K_0 \text{ nearest neighbours of } B \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

where K_0 is an integer constant and the only parameter of the model, and k is a normalization constant that corresponds to the number of individuals that possesses A within their K_0 nearest neighbours.

As in the previous analyzed model, here the model parameter k has to be such that percolation is ensured; note this time the percolated network is directed as shown in Fig.1h. This means that K_0 has to be $K_0 > K_*$ as shown in Fig. 2c. As before, the value of K_0 that best describes the experimental data is obtained by MLE. This model, as indicated by the KS-test [D-statistics, $D = 0.23 < D_{\text{critical},0.05} = 0.33$] is also consistent with the experimental data; the resulting simulated CDD is shown in Fig. 2d. The idea that there is a hard threshold that regulates the probability that B selects A as interaction-neighbour, as suggested by eqs. 2 and 3, seems, from a biological point of view, unrealistic. In the following, we explore the possibility, arguably more realistic, that $W(A, B)$ is given by a smooth, continuous decreasing function of either the Euclidean distance or distance rank between A and B characterized, again, by only one parameter. Using a smooth, continuous function for $W(A, B)$ has an important conceptual consequence. Under this assumption, any individual can potentially become an interaction neighbour of another group member, though this may occur with a very small probability if the two individuals are located far away in the group. A priori, any decreasing functional form, characterized by a single parameter, could be tested. For simplicity, we limit the discussion to an exponential distribution, which is widespread in statistics, easy to interpret, and is characterized by one parameter, the so-called rate parameter. Assuming that $W(A, B)$ depends on Euclidean distance d_{MBA} from B to A , we define the probability as:

$$W(A, B) = \frac{e^{-\frac{d_{MBA}}{R_c}}}{Z} \quad (4)$$

where R_C is a constant, and the only parameter of the model, and Z defined by

$$Z = \sum_{B \neq A} e^{-\frac{d_{MBA}}{R_C}}$$

ensures that $W(A, B)$ is properly normalized. We used MLE to obtain the value of R_C that best describes the data. This model, as indicated by the *KS*-test [*D*-statistics, $D = 0.24 < D_{\text{critical},0.05} = 0.33$] is also consistent with the experimental data; the resulting simulated CDD is shown in Fig. 2e. If instead we assume that $W(A, B)$ depends on the distance rank between A and B , the same concept leads to a probability of the form:

$$W(A, B) = \frac{e^{-\frac{d_{TBA}}{K_C}}}{Z} \quad (5)$$

where d_{TBA} is the distance rank of A with respect to B (i.e. whether A is the first, second, third, etc., nearest neighbour from B), K_C is a constant, and unique parameter of the model, and Z defined as

$$Z = \sum_{B \neq A} e^{-\frac{d_{TBA}}{T_C}}$$

ensures that $W(A, B)$ is properly normalized. The *KS*-test [*D*-statistics, $D = 0.18 < D_{\text{critical},0.05} = 0.33$] applied to this model using the value of K_C that best describes the data (obtained by MLE) indicates that eq. 5 is also consistent with the data.

Since the selection models inspired in the standard metric and topological neighbour models (eqs. 2 and 3) are consistent with the experimental data, as well as the models that assume that $W(A, B)$ is given by a smooth, continuous decreasing function of either the distance or distance rank between A

and B (eqs. 4 and 5), we use the negative log-likelihood – defined as $l = -\log(L)$, where L is the likelihood -- to identify the model that best describes the data. Note that since we are comparing models with the same number of parameters, this is equivalent to using the Akaike model selection criterion. The obtained negative log-likelihood values are: i) for the selection model inspired on the standard metric model, 60.72, ii) for the one inspired on the standard topological model, 61.15, iii) for the exponential metric model, 57.99, and iv) for the exponential topologic model, 54.67. This means that the models defined by eqs. 4 and 5 perform better than the one for given by eqs. 2 and 3. It is worth indicating that even though it is tempting to use the negative log-likelihood to select the model given by eq. 5 over the model defined by Eq. 4, a *KS*-test over simulated data from these two models indicate they are statistically indistinguishable [*D*-statistics, $D = 0.1016 < D_{\text{critical}, 0.05} = 0.1025$] when using the optimal values R_c and K_c obtained by MLE. In short, we cannot distinguish between the model given by eq. 4 and the one by eq. 5.

Combining Distance and Distance Rank models – a 2 parameters model

Since we cannot distinguish between the models defined eqs. 4 and 5, we combine both models in an attempt to measure the relative importance in $W(A, B)$ of Euclidean distance vs. distance rank. The combined model reads:

$$W(A, B) = \frac{e^{-\frac{d_{MA}}{R_c}} \times e^{-\frac{d_{TB}}{K_c}}}{Z} \quad (6)$$

where

$$Z = \sum_{B \neq A} e^{-\frac{d_{MA}}{R_c}} \times e^{-\frac{d_{TB}}{K_c}}$$

and the two parameters of the model are R_c and K_c . These parameters are again obtained by MLE. Not surprisingly, this model, with two parameters, is also consistent with the data [D -statistics, $D = 0.18 < D_{\text{critical}, 0.05} = 0.33$]. We insist that this model with 2 parameters is not introduced as an alternative to the one-parameter models, but as a tool to assess the relative weight in $W(A, B)$ of Euclidean distance vs. distance rank. Here, we are interested in the values of R_c and K_c , namely, $R_c = 83$ m and $K_c = 14$. Considering that $N = 32$ and the diagonal of the squared arena is 70.7 m, this suggests that $W(A, B)$ is more sensitive to difference in distance rank than in Euclidean distance. Nevertheless, the observation that a model that combines both, distance and distance rank, is consistent with the empirical data points towards more complex processes being involved in the interaction-neighbour selection, where both, the distance rank and the distance between two individuals play a role.

DISCUSSION

Information spreading and processing within animal groups – that plays a key role in complex processes such as collective motion and predator detection – strongly depends on the properties of the underlying interaction network. In the absence of a communication system (e.g. vocalization), information spreading takes the form of a propagation of behavioural change. Despite the relevance of identifying how interaction partners are selected to understand the dynamics of animal groups, there are only few examples of experiments specially designed to decipher the underlying interaction network (Jiang et al., 2017; Rosenthal et al., 2015; Strandburg-Peshkin et al., 2013). In most experimental conditions, behavioural changes are difficult to identify, and the presence of multiple sources (individuals) participating in the behavioural-change propagation makes it difficult to determine with certainty to whom an individual is reacting to. There exist, however, few notable experimental studies

of animal groups on the move, where influential neighbours were identified (Ballerini et al., 2008; Herbert-Read et al., 2011; Jiang et al., 2017; Rosenthal et al., 2015; Strandburg-Peshkin et al., 2013), but only indirectly, after a data analysis treatment allowed establishing correlations in velocity changes among group members.

Here, we investigated how interaction neighbours are selected, and focused on the early stage of a collective displacement, when the interaction network is first established. The experiments were specifically designed to identify the selection process of interaction neighbours. In these experiments, all group members share, initially, the same behavioural state, except for a single individual. This experimental scenario was achieved by introducing a trained individual that we were able to activate remotely, within a group of naive individuals. This allowed us to initiate experiments only once all individuals were in the desired initial behavioural state (a non-motion state). After setting in motion the trained individual, we monitored the behavioural state of all group members, and identified the first individual that followed (and thus that chose as interaction neighbour) the trained individual. Finally, by focusing on the first follower, we avoided the ambiguity that emerges by the presence of multiple activation sources in the system (the who-activates-whom problem) that takes place as the behavioural change propagates through the group. This allowed us to collect reliable experimental data to investigate the interaction-neighbour selection process, which we characterized with $W(A, B)$: the probability that B chooses A as interaction neighbour.

In models that aim at explaining collective displacements (e.g. Calovi et al., 2018; Heras et al., 2019; Herbert-Read et al., 2011; Katz et al., 2011; Tang et al., 2020; Torney et al., 2018), it has been proposed the use of neighbourhood rules – i.e. rules that allow to determine the neighbours of any given individual at a given time – that include, in the equation of motion of the individuals, a continuous, distant-dependent weight associated to each neighbour. It is worth stressing that here we are not computing these weights nor modeling collective displacements. The difference is that the present study focuses on the cognitive selection process used by the individuals to choose interaction

neighbours, and not on the weights assigned to neighbours during a collective displacement. This implies the assumption that the individual behaviour can be dissected in a series of simple cognitive tasks: first, select your interaction neighbours – the focus on the current study – then execute an action (e.g. choose a direction of motion), in which the selected interaction neighbours may be weighted differently. Assuming the existence of a cognitive selection process of interaction neighbours has a strong impact on the concept of neighbourhood. For instance, the interaction neighbours of an individual are not necessarily determined by the instantaneous position of the individuals, as assumed in most collective displacement models, but are likely to depend on the recent “history” of the individual and the system. For a situation where all individuals are initially at fixed position, as analyzed in this study, it is reasonable to assume that the selection process is fundamentally dependent on the relative position of the individuals. And indeed we found that $W(A, B)$ is modulated by the relative distance between the individuals, observing that the best estimations are obtained using exponential decaying functions of either the distance or distance rank. And while $W(A, B)$ seems to be more sensitive to the distance rank than to the (Euclidean) distance, it is important to stress that the distance dependency is always present and not negligible.

In summary, these results, obtained by following a protocol that allows assessing $W(A, B)$ in a direct and reliable manner, provide unique, valuable information on how individuals select interaction neighbours within groups. Knowledge on the functional form of $W(A, B)$ is key to understand the spreading of behavioural changes, and thus information, in animal groups. It is likely that other animal systems operate using a similar $W(A, B)$, while the outlined protocol should be applicable to assess the structure of $W(A, B)$ in other biological contexts.

Data statement

The data that support the plots within this paper and other findings of this study are available from the corresponding author upon request.

Acknowledgments

We thank the staff of the Domaine du Merle and M.-H. Pillot for support during experiments, S. Toulet, R. Violleau for their contribution to the data collection, and P. Arrufat for technical help. S. Toulet contributed to data analyses. This work was supported by one CNRS PEPS grant Physique Théorique et ses Interfaces and one CNRS PEPS grant BIO-MATHS-INFO. Two anonymous reviewers provide helpful feedback. Authors have no conflict of interest to declare. HG was supported by a PhD grant of the French Ministry of Superior Education and Research

References

- Amé, J.-M., Halloy, J., Rivault, C., Detrain, C., & Deneubourg, J.-L. (2006). Collegial decision making based on social amplification leads to optimal group formation. *Proceedings of the National Academy of Sciences*, *103*, 5835–5840. www.pnas.org/cgi/doi/10.1073/pnas.0507877103
- Aoki, I. (1982). A Simulation Study on the Schooling Mechanism in Fish. *Bulletin of the Japanese Society of Scientific Fisheries*, *48*, 1081–1088. <https://doi.org/10.2331/suisan.48.1081>
- Attanasi, A., Cavagna, A., Del Castello, L., Giardina, I., Grigera, T. S., Jelic, A., Melillo, S., Parisi, L., Pohl, O., Shen, E., & Viale, M. (2014). Information transfer and behavioural inertia in starling flocks. *Nature Physics*, *10*, 691–696. <https://doi.org/10.1038/nphys3035>
- Ballerini, M., Cabibbo, N., Candelier, R., Cavagna, A., Cisbani, E., Giardina, I., Lecomte, V., Orlandi, A., Parisi, G., & Procaccini, A. (2008). Interaction ruling animal collective behavior depends on

topological rather than metric distance: Evidence from a field study. *Proceedings of the National Academy of Sciences*, *105*, 1232-1237. www.pnas.org/cgi/doi/10.1073/pnas.0711437105

Barberis, L. & Peruani, F. (2016). Large-scale patterns in a minimal cognitive flocking model: incidental leaders, nematic patterns, and aggregates. *Physical Review Letters*, *117*, 248001. <https://doi.org/10.1103/PhysRevLett.117.248001>

Bloch, G., Herzog, E. D., Levine, J. D., & Schwartz, W. J. (2013). Socially synchronized circadian oscillators. *Proceedings of the Royal Society of London Series B-Biological Sciences*, *280*, 20130035. <http://dx.doi.org/10.1098/rspb.2013.0035>

Buhl, J., Sumpter, D. J. T., Couzin, I. D., Hale, J. J., Despland, E., Miller, E. R., & Simpson, S. J. (2006). From disorder to order in marching locusts. *Science*, *312*, 1402–1406. <https://doi.org/10.1126/science.1125142>

Calovi, D. S., Litchinko, A., Lecheval, V., Lopez, U., Escudero, A. P., Chaté, H., Sire, C., & Theraulaz, G. (2018). Disentangling and modeling interactions in fish with burst-and-coast swimming reveal distinct alignment and attraction behaviors. *PLoS Computational Biology*, *14*, e1005933. <https://doi.org/10.1371/journal.pcbi.1005933>

Camazine, S., Deneubourg, J.-L., Franks, N. R., Sneyd, J., Theraulaz, G., & Bonabeau, E. (2001). *Self-Organization in Biological Systems*. Princeton & Oxford: Princeton University Press.

Casellas, E., Gautrais, J., Fournier, R., Blanco, S., Combe, M., Fourcassié, V., Theraulaz, G., & Jost, C. (2008). From individual to collective displacements in heterogeneous environments. *Journal of Theoretical Biology*, *250*, 424–434. <https://doi.org/10.1016/j.jtbi.2007.10.011>

Chauvin, R. (1982). *Les sociétés animales*. Paris : Presses Universitaires de France.

Chepizhko, O., Altmann, E., & Peruani, F. (2013). Optimal Noise Maximizes Collective Motion in Heterogeneous Media. *Physical Review Letters*, *110*, 238101. <https://doi.org/10.1103/PhysRevLett.110.238101>

- Conradt, L., Krause, J., Couzin, I. D., & Roper, T. J. (2009). "Leading according to need" in self-organizing groups. *The American Naturalist*, *173*, 304–312. <https://doi.org/10.1086/596532>
- Couzin, I. D., Krause, J., James, R., Ruxton, G. D., & Franks, N. R. (2002). Collective memory and spatial sorting in animal groups. *Journal of Theoretical Biology*, *218*, 1–12. <https://doi.org/10.1006/jtbi.2002.3065>
- Doligez, B., Danchin, É., & Clobert, J. (2002). Public information and breeding habitat selection in a wild bird population. *Science*, *297*, 1168–1170. <https://doi.org/10.1126/science.1072838>
- Fischhoff, I. R., Sundaresan, S. R., Cordingley, J., Larkin, H. M., Sellier, M.-J., & Rubenstein, D. I. (2007). Social relationships and reproductive state influence leadership roles in movements of plains zebra, *Equus burchellii*. *Animal Behaviour*, *73*, 825–831. <https://doi.org/10.1016/j.anbehav.2006.10.012>
- Ginelli, F., Peruani, F., Pillot, M.-H., Chaté, H., Theraulaz, G., & Bon, R. (2015). Intermittent collective dynamics emerge from conflicting imperatives in sheep herds. *Proceedings of the National Academy of Sciences*, *112*, 12729–12734. <https://doi.org/10.1073/pnas.1503749112>
- Giraldeau, L.-A. & Caraco, T. (2000). *Social Foraging Theory*. Princeton: Princeton University Press.
- Grassé, P.-P. (1946). Sociétés animales et effet de groupe. *Experientia*, *2*, 77–82.
- Grassé, P.-P. (1968). *L'effet de groupe chez les animaux*. Paris : Éditions du Centre National de la Recherche Scientifique.
- Heras, F. J., Romero-Ferrero, F., Hinz, R. C., & de Polavieja, G. G. (2019). Deep attention networks reveal the rules of collective motion in zebrafish. *PLoS Computational Biology*, *15*, e1007354. <https://doi.org/10.1371/journal.pcbi.1007354>
- Herbert-Read, J. E., Perna, A., Mann, R. P., Schaerf, T. M., Sumpter, D. J. T., & Ward, A. J. W. (2011). Inferring the rules of interaction of shoaling fish. *Proceedings of the National Academy of Sciences*, *108*, 18726–18731. <https://doi.org/10.1073/pnas.1109355108>

- James, F. (2006). *Statistical methods in experimental physics*. London: World Scientific Publishing Company.
- Jiang, L., Giuggioli, L., Perna, A., Escobedo, R., Lecheval, V., Sire, C., Han, Z., & Theraulaz, G. (2017). Identifying influential neighbours in animal flocking. *PLoS Computational Biology*, *13*, e1005822. <https://doi.org/10.1371/journal.pcbi.1005822>
- Jolles, J. W., Mazué, G. P., Davidson, J., Behrmann-Godel, J., & Couzin, I. D. (2020). *Schistocephalus* parasite infection alters sticklebacks' movement ability and thereby shapes social interactions. *Scientific Reports*, *10*, 1–11. <https://doi.org/10.1038/s41598-020-69057-0>
- Katz, Y., Tunstrøm, K., Ioannou, C. C., Huepe, C., & Couzin, I. D. (2011). Inferring the structure and dynamics of interactions in schooling fish. *Proceedings of the National Academy of Sciences*, *108*, 18720–18725. <https://doi.org/10.1073/pnas.1107583108>
- Kim, H. (2014). Don't Count The Number of Friends When You Are Spreading Information in Social Networks. Proceedings of the 8th International Conference on Ubiquitous Information Management and Communication, pages 1–7.
- Kunz, H. & Hemelrijk, C. K. (2012). Simulations of the social organization of large schools of fish whose perception is obstructed. *Applied Animal Behaviour Science*, *138*, 142–151. <https://doi.org/10.1016/j.applanim.2012.02.002>
- Lesne, P. & Jeanson, R. (2015). Influence of social interactions on the response to social cues in spiderlings. *Behavioural Processes*, *111*, 60–66. <https://doi.org/10.1016/j.beproc.2014.11.018>
- Lihoreau, M. & Rivault, C. (2011). Local Enhancement Promotes Cockroach Feeding Aggregations. *PloS One*, *6*, e22048. <https://doi.org/10.1371/journal.pone.0022048>
- Lima, S. L. (1995). Collective detection of predatory attack by social foragers: fraught with ambiguity? *Animal Behaviour*, *50*, 1097–1108. [https://doi.org/10.1016/0003-3472\(95\)80109-X](https://doi.org/10.1016/0003-3472(95)80109-X)

- Parrish, J. K. & Edelstein-Keshet, L. (1999). Complexity, pattern, and evolutionary trade-offs in animal aggregation. *Science*, 284, 99–101. <https://doi.org/10.1126/science.284.5411.99>
- Pillot, M.-H., Gautrais, J., Arrufat, P., Couzin, I. D., Bon, R., & Deneubourg, J.-L. (2011). Scalable Rules for coherent Group Motion in a Gregarious Vertebrate. *PloS One*, 6, e14487. <https://doi.org/10.1371/journal.pone.0014487>
- Pillot, M.-H., Gautrais, J., Gouello, J., Michelena, P., Sibbald, A. M., & Bon, R. (2010). Moving together: Incidental leaders and naive followers. *Behavioural Processes*, 83, 7–7. <https://doi.org/10.1016/j.beproc.2009.11.006>
- Reebs, S. G. (2001). Influence of body size on leadership in shoals of golden shiners, *Notemigonus crysoleucas*. *Behaviour*, 138, 797–809. <https://doi.org/10.1163/156853901753172656>
- Reuter, H. & Breckling, B. (1994). Self-organization of fish schools: an object-oriented model. *Ecological Modelling*, 75-76, 147–159.
- Rosenthal, S. B., Twomey, C. R., Hartnett, A. T., Wu, H. S., & Couzin, I. D. (2015). Revealing the hidden networks of interaction in mobile animal groups allows prediction of complex behavioral contagion. *Proceedings of the National Academy of Sciences*, 112, 4690–4695. <https://doi.org/10.1073/pnas.1420068112>
- Strandburg-Peshkin, A., Twomey, C. R., Bode, N. W. F., Kao, A. B., Katz, Y., Ioannou, C. C., Rosenthal, S. B., Torney, C. J., Wu, H. S., Levin, S. A., & Couzin, I. D. (2013). Visual sensory networks and effective information transfer in animal groups. *Current Biology*, 23, R709–R711. <https://doi.org/10.1016/j.cub.2013.07.059>
- Sueur, C., Deneubourg, J.-L., & Petit, O. (2012). From Social Network (Centralized vs. Decentralized) to Collective Decision-Making (Unshared vs. Shared Consensus). *PloS One*, 7, e32566. <https://doi.org/10.1371/journal.pone.0032566>

- Sueur, C., Petit, O., & Deneubourg, J.-L. (2009). Selective mimetism at departure in collective movements of *Macaca tonkeana*: an experimental and theoretical approach. *Animal Behaviour*, *78*, 1087–1095. <https://doi.org/10.1016/j.anbehav.2009.07.029>
- Sueur, C., Petit, O., & Deneubourg, J.-L. (2010). Short-term group fission processes in macaques: a social networking approach. *Journal of Experimental Biology*, *213*, 1338–1346. <https://doi.org/10.1242/jeb.039016>
- Sumpter, D. J. T. (2006). The principles of collective animal behaviour. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *361*, 5–22. <https://doi.org/10.1098/rstb.2005.1733>
- Tang, W., Davidson, J. D., Zhang, G., Conen, K. E., Fang, J., Serluca, F., Li, J., Xiong, X., Coble, M., Tsai, T., et al. (2020). Genetic control of collective behavior in zebrafish. *iScience*, *23*, 100942. <https://doi.org/10.1016/j.isci.2020.100942>
- Tinbergen, N. (1963). On aims and methods of ethology. *Zeitschrift für Tierpsychologie*, *20*, 410-433. <https://doi.org/10.1111/j.1439-0310.1963.tb01161.x>
- Torney, C. J., Lamont, M., Debell, L., Angohiatok, R. J., Leclerc, L.-M., & Berdahl, A. M. (2018). Inferring the rules of social interaction in migrating caribou. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *373*, 20170385. <http://dx.doi.org/10.1098/rstb.2017.0385>
- Toulet, S., Gautrais, J., Bon, R., & Peruani, F. (2015). Imitation combined with a characteristic stimulus duration results in robust collective decision-making. *PloS One*, *10(10)*, e0140188. □ <https://doi.org/10.1371/journal.pone.0140188>
- Valone, T. J. & Templeton, J. J. (2002). Public information for the assessment of quality: a widespread social phenomenon. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *357*, 1549–1557. <https://doi.org/10.1098/rstb.2002.1064>

- Vicsek, T., Czirók, A., Ben-Jacob, E., Cohen, I., & Shochet, O. (1995). Novel type of phase transition in a system of self-driven particles. *Physical Review Letters*, 75, 1226–1229. <https://doi.org/10.1103/PhysRevLett.75.1226>
- Viscido, S. V., Miller, M., & Wethey, D. S. (2002). The dilemma of the selfish herd: the search for a realistic movement rule. *Journal of Theoretical Biology*, 217, 183–194. <https://doi.org/10.1006/jtbi.2002.3025>
- Viscido, S. V., Parrish, J. K., & Grünbaum, D. (2005). The effect of population size and number of influential neighbours on the emergent properties of fish schools. *Ecological Modelling*, 183, 347–363. <https://doi.org/10.1016/j.ecolmodel.2004.08.019>
- Wyatt, T. D. (2003). *Pheromones and Animal Behaviour. Communication by Smell and Taste*. Cambridge: Cambridge University Press,.

Figure captions

Figure 1. The interaction-neighbour selection probability $W(A, B)$. Histogram of the Euclidian distance (a) and distance rank (b) between the trained individual (initiator) and the first group member to respond to the initiator departure. The first follower tend to be near the initiator in both, Euclidian distance and distance rank. Panels (c) and (d) illustrate the symmetry of the selection probability $W(A, B)$ when based on Euclidean distance. On the other hand, when $W(A, B)$ is based on a distance rank, tends to be asymmetric. Resulting interaction networks strongly depends on whether $W(A, B)$ was based on a metric distance or distance rank, as illustrated in panels (g) and (h).

Figure 2. Comparison of different interaction-neighbour selection models with experimental sheep-herd data. The experimental cumulative distributions of first follower distances (plain black curves) and the theoretical models: (a) the global perception model, (d) the model inspired on the standard metric neighbourhood model (for short, standard metric), eq. 2, the one inspired on the standard distance rank neighbourhood model (for short, standard topologic) , eq. 3, (e) the metric model defined by eq. 4 (Exponential metric) and the topological one given by eq. 5 (exponential topologic), and (f) the model combining both, distance and distance rank, eq. 6. Probability that networks are percolated as a function of the values of the model parameter for (b) the standard metric model, eq. 2 ($R = 11$), and (c) distance rank model, eq. 5 ($K = 8$).



