



**HAL**  
open science

## A data-driven method for reconstructing and modelling social interactions in moving animal groups

R. Escobedo, V. Lecheval, V. Papaspyros, F. Bonnet, F. Mondada, Clément Sire, Guy Theraulaz

### ► To cite this version:

R. Escobedo, V. Lecheval, V. Papaspyros, F. Bonnet, F. Mondada, et al.. A data-driven method for reconstructing and modelling social interactions in moving animal groups. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 2020, 375 (1807), pp.20190380. 10.1098/rstb.2019.0380 . hal-03363025

**HAL Id: hal-03363025**

**<https://hal.science/hal-03363025>**

Submitted on 2 Oct 2021

**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.

# A data-driven method for reconstructing and modelling social interactions in moving animal groups

R. Escobedo<sup>1</sup>, V. Lecheval<sup>2</sup>, V. Pappaspyros<sup>3</sup>, F. Bonnet<sup>3</sup>, F. Mondada<sup>3</sup>, C. Sire<sup>4</sup>, G. Theraulaz<sup>1,5,\*</sup>

<sup>1</sup> Centre de Recherches sur la Cognition Animale, Centre de Biologie Intégrative (CBI), Centre National de la Recherche Scientifique (CNRS) & Université de Toulouse – Paul Sabatier, 31062 Toulouse, France

<sup>2</sup> Department of Biology, University of York, Heslington YO10 5DD York, United Kingdom

<sup>3</sup> MOBOTS group, Biorobotics laboratory, École Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland

<sup>4</sup> Laboratoire de Physique Théorique, CNRS & Université de Toulouse – Paul Sabatier, 31062 Toulouse, France

<sup>5</sup> Centre for Ecological Sciences, Indian Institute of Science, Bengaluru, India

\*guy.theraulaz@univ-tlse3.fr

– 02 February 2020 –

**Abstract** Group-living organisms that collectively migrate range from cells and bacteria to human crowds, and include swarms of insects, schools of fish, and flocks of birds or ungulates. Unveiling the behavioural and cognitive mechanisms by which these groups coordinate their movements is a challenging task. These mechanisms take place at the individual scale and can be described as a combination of interactions between individuals and interactions between these individuals and the physical obstacles in the environment. Thanks to the development of novel tracking techniques that provide large and accurate data sets, the main characteristics of individual and collective behavioural patterns can be quantified with an unprecedented level of precision. However, in a large number of studies, social interactions are usually described by force map methods that only have a limited capacity of explanation and prediction, being rarely suitable for a direct implementation in a concise and explicit mathematical model. Here, we present a general method to extract the interactions between individuals that are involved in the coordination of collective movements in groups of organisms. We then apply this method to characterize social interactions in two species of shoaling fish, the rummy-nose tetra (*Hemigrammus rhodostomus*) and the zebrafish (*Danio rerio*), which both present a burst-and-coast motion. From the detailed quantitative description of individual-level interactions, it is thus possible to develop a quantitative model of the emergent dynamics observed at the group-level, whose predictions can be checked against experimental results. This method can be applied to a wide range of biological and social systems.

## 1 Introduction

The identification and characterization of interactions between the constituent elements of a living system is a major challenge for understanding its dynamic and adaptive properties [1, 2]. In recent years, this issue is also at the heart of research conducted in the field of collective behaviour in animal societies [3, 4]. However, the identification from field data of the interaction rules between individuals in species whose level of cog-

nitive complexity can be quite high remains problematic. Indeed, the way in which individuals interact is strongly influenced and modulated by the physical characteristics of the environment in which the organisms live, such as temperature, humidity, brightness, or even by the presence of air currents (see for instance [5, 6] in social insects). The situation is quite different in the laboratory, where conditions can be precisely controlled and monitored. Moreover, new tracking techniques make it possible to record the behaviour of individuals alone or in groups for relatively long periods of time [7, 8, 9, 10, 11]. Using large sets of tracking data, one can then reconstruct and model the social interactions between two individuals of the same species, and between them and the obstacles present in their environment [12, 13].

Explaining collective behaviour in groups of organisms consists in describing the mechanisms by which the behaviours of an individual is influenced by the behaviour of the other group members that are present in its neighbourhood [14, 15]. This social influence is often assumed to result from the additive combination of pairwise interactions of the individual with part or all the members of the group. Determining how these pairwise interactions may be combined, and which neighbours must be taken into account, is a central problem in the study of collective behaviour [16].

The behavioural response of an individual is the set of its successive positions during a given period of time, usually at some discrete time steps. From these data, it is possible to draw the trajectories of all the individuals in a group and calculate their instantaneous velocity and acceleration, their distance and angle of incidence to obstacles (*e.g.*, the wall of an experimental tank), their heading, as well as relative quantities such as the distance between individuals, the angle of their relative position, and group quantities such as cohesion and heading polarization. These measures can reveal individual behavioural patterns such as the average velocity or the frequency of heading changes close to obstacles, and also collective behavioural patterns such as the level of cohesion and polarization of the group. Thus, the analysis of collective behaviour consists in measuring behavioural changes at the individual

scale that likely result from social interactions, and to associate these measures with the relative state of the individuals involved in these interactions [12]. The relative state of an individual  $j$  with respect to a focal individual  $i$  is determined by the distance  $d_{ij}$  between  $i$  and  $j$ , its relative velocity  $v_{ij}$ , its angular position with respect to  $i$ ,  $\psi_{ij}$  (viewing angle), and its relative heading  $\phi_{ij}$  (see Fig. 1). The behavioural changes are precisely given by the variations of an individual’s position and velocity, or, equivalently, by the position, speed and heading variations.

In fish that have a burst-and-coast swimming mode, the behavioural changes of an individual correspond to significant variations of its heading that occur exactly at the onset of the acceleration phase (*i.e.*, the bursts). These discrete behavioural decisions are called “kicks” [12, 17]. Other quantities such as the intensity of the acceleration in the direction perpendicular to the direction of motion, or simply the turning direction (right or left), can be used to detect behavioural changes. The task is thus to relate the heading variation of a focal fish  $\delta\phi_i$  with its state variables, that is, to find a function  $\delta\phi_i(d, v, \psi, \phi)$ .

In this article, we first show that force maps, that are widely used to describe the effects of social interactions on the behaviour of individuals, have important limitations when it comes to describe these interactions. These limitations result mostly from the limited number of variables that force maps can handle, the difficulty of identifying intermediate contributions to behavioural patterns, and the difficulty of distinguishing the effects of state variables from constitutive parameters. We then describe in detail a method to analyse behavioural data obtained from digitized individual trajectories. The method allows us (1) to quantify the social interactions between two individuals and describe how the intensities of these interactions vary as a function of the state variables of the individuals, and, (2) to reconstruct the interaction functions and to derive an explicit and concise mathematical model reproducing the observed behaviours. Finally, we apply this method to the analysis of the social interactions in two species of fish that both have a burst-and-coast type of swimming and that are characterized by very different levels of coordination when swimming in groups. The reconstruction of interaction rules allows to understand the origin of the differences in the level of coordination, and to predict in which experimental conditions other behavioural differences can arise.

## 2 Use and limitations of force maps to infer social interactions

A first way to infer social interactions between individuals directly from experimental data consists in using the force-map technique [13, 18]. This technique has been

used, for instance, to estimate from experiments performed with two fish the effective turning and speeding forces experienced by an individual, once the relevant variables on which they may depend have been chosen [11, 18, 19]. This is a simple way to visualise the strength and direction of behavioural changes. However, force maps have strong limitations that can induce profound misunderstandings.

### 2.1 Visualisation of social interactions with force maps

Force maps are planar representations of two-dimensional functions of the form  $f(x, y)$  where the variation of the value of the function is represented by a colour gradient. Generating and interpreting force maps is easy and this explains their success for inferring interactions between moving groups of individuals.

A 2D function  $f(x, y)$  given by a data set is a sequence of  $N$  triplets  $(x^n, y^n, f^n)$ , where the index  $n$  denotes for example the instant of time  $t^n$ ,  $n = 1, \dots, N$ . To build a force map of this function, the  $(x, y)$ -space is discretised in  $I \times J$  boxes of the form  $[\hat{x}_i, \hat{x}_{i+1}] \times [\hat{y}_j, \hat{y}_{j+1}]$ , where the nodes  $\hat{x}_i$  and  $\hat{y}_j$  are given by

$$\begin{aligned}\hat{x}_i &= \hat{x}_{\min} + (i - 1)(\hat{x}_{\max} - \hat{x}_{\min})/I, \quad i = 1, \dots, I + 1, \\ \hat{y}_j &= \hat{y}_{\min} + (j - 1)(\hat{y}_{\max} - \hat{y}_{\min})/J, \quad j = 1, \dots, J + 1,\end{aligned}$$

and the data  $(x^n, y^n, f^n)$  is placed in the  $ij$ -box such that  $x^n$  is in  $[\hat{x}_i, \hat{x}_{i+1}]$  and  $y^n$  is in  $[\hat{y}_j, \hat{y}_{j+1}]$ . Then, the number of data  $\epsilon_{ij}$  in the  $ij$ -box and the mean value  $f_{ij}$  of the values of  $f^n$  that fell in the  $ij$ -box are calculated. The value  $f_{ij}$  is then considered as the value of  $f(x, y)$  at the middle point of the  $ij$ -box,  $x_i = (\hat{x}_i + \hat{x}_{i+1})/2$ ,  $y_i = (\hat{y}_i + \hat{y}_{i+1})/2$ . The resulting points  $(x_i, y_j, f_{ij})$  are then represented in a colour surface, after optional interpolation with, for instance, multilevel B-splines [20].

Figs. 2 and 3 show the force maps of the heading variation  $\delta\phi$  of an individual fish performing a burst-and-coast type of swimming, as a function of different variables when there is another conspecific in a circular tank. In the case of *Hemigrammus rhodostomus* (Fig. 2), we used a tank of radius  $R = 0.25$  m, about 8.3 times the body length (BL) of the fish, and in the case of *Danio rerio* (Fig. 3), a tank of radius  $R_Z = 0.29$  m, about 6.4 BL. In both Figs. 2 and 3, panels A show the intensity of  $\delta\phi$  as a function of  $d_{ij}$ , the distance between fish, and  $\psi_{ij}$ , the angle with which fish  $i$  perceives fish  $j$ , respectively, and panels B show  $\delta\phi$  as a function of  $d_{ij}$  and  $\phi_{ij}$ , the heading difference between fish.

These force maps provide some information about the individual behaviour of fish. In *H. rhodostomus*, Fig. 2A shows that the focal fish tends to turn towards its neighbour, to the left (resp. right) when the neighbour is on the left (resp. right), except when the neighbour is very close. In that case, the behaviour is quite complex, the fish performs small angular changes of amplitude  $\approx 30$ – $60^\circ$ , probably due to collision avoidance

manoeuvres. When the neighbour is further away from the focal fish, it maintains its heading (*i.e.*, the white circular region at  $d_{ij} \approx 1-2\text{BL}$ ). From this force map, one could conclude that a fish is attracted by its neighbour when it is beyond a distance of about  $2\text{BL}$ , and repulsed when its neighbour is too close ( $d_{ij} < 1\text{BL}$ ). The force map in Fig. 2B shows that the focal fish turns left (resp. right) when the relative heading of the neighbour is shifted to the left (resp. right). The larger the heading difference, the stronger the turn: the colour intensity increases as  $|\phi_{ij}|$  grows from  $0$  to  $120^\circ$ . When fish swim in more or less opposite directions, the intensity of the heading change is small. This force map reveals that a fish tends to align with its neighbour. In *D. rerio*, Fig. 3A shows that the focal fish turns towards its neighbour when they are close to each other ( $d_{ij} \approx 1-2\text{BL}$ ) or when the neighbour is located behind the focal fish ( $|\psi_{ij}| > 90^\circ$ , whatever the distance). When the neighbour is at  $d_{ij} \approx 2-3\text{BL}$  in front of the focal fish ( $|\psi_{ij}| < 60^\circ$ ), the focal fish turns away from its neighbour, as well as when its neighbour is very close to it ( $d_{ij} < 0.5\text{BL}$ ). The reaction to neighbour's heading is less intense than in *H. rhodostomus*, as shown by the wide white regions in Fig. 3B. When fish are beyond  $3\text{BL}$  from each other, the focal fish turns to adopt the same orientation of its neighbour. At short distances, the behaviour is more complex, with changes of smaller size than those observed in *H. rhodostomus*.

The visualisation of the data by means of force maps therefore suggests the presence of two distinct types of interaction: an attraction interaction, which leads a fish to turn towards its neighbour to get closer to it, and an alignment interaction, that leads a fish to turn so as to adopt the same heading than its neighbour. However, nothing can be deduced from these colour maps about what happens when both contributions to heading variation have different signs. Thus, would a fish turn right or left when its neighbour is on its left side? Attraction alone would induce the focal fish to turn left. However, if the relative heading of the neighbour is turned to the right, alignment alone would induce the focal fish to turn right. This difficulty comes from the fact that a function ( $\delta\phi$ ) that depends on three variables ( $d_{ij}$ ,  $\psi_{ij}$ , and  $\phi_{ij}$ ) cannot be represented in 3D. To overcome this limitation, some authors use a kind of force map where the relative position of a fish with respect to a focal fish is decomposed in the left-right (LR) and front-back (FB) distances [13]. In Fig. 2A,  $(d_{ij}, \psi_{ij})$  are the polar coordinates of fish  $j$  in the system of reference centred on fish  $i$  pointing north. This is a continuous system of reference in which all the relative positions of fish can be represented. Instead, the LR and FB distances are projections of the relative position of a neighbour with respect to the focal fish on the  $d_{ij}$ -axis, where all the points of Fig. 2A that are in the left semicircle of radius  $1\text{BL}$  are averaged in a single point where  $d_{ij}^{\text{LR}} = -1\text{BL}$  (where  $d_{ij}^{\text{LR}}$  is the LR distance of fish  $j$  with respect to fish  $i$ ), because these points are at  $1\text{BL}$  to the left of the focal fish. Then, the third variable  $\phi_{ij}$  is used to ex-

pand this averaged point on a vertical line with different values of  $\phi_{ij}$ , in a system of reference with coordinates  $(d_{ij}^{\text{LR}}, \phi_{ij})$ , giving rise to Fig. 2C. Similarly, the upper semicircles of Fig. 2A are averaged on the  $d_{ij}^{\text{FB}}$ -axis in Fig. 2D.

Force maps in panels C and D of Figs. 2 and 3 provide additional information about individual fish behaviour. In *H. rhodostomus*, Fig. 3C shows that turning direction is homogeneously distributed in the upper and lower half-planes, meaning that the focal fish turns to adopt the heading of its neighbour almost independently of the LR distance separating them, although the turning intensity is larger when the neighbour is far from the focal fish and perpendicular to it (*i.e.*, the regions of highly intense colour at  $|d_{ij}^{\text{LR}}| > 2\text{BL}$  and  $\phi_{ij} \approx \pm 90^\circ$ ). In the white horizontal region, the focal fish maintains its heading when it is aligned with its neighbour ( $|\phi_{ij}| < 10^\circ$ ), whatever the horizontal distance between them. Fig. 2D exhibits two large regions homogeneous in colour, showing that the focal fish turns almost always to adopt the heading direction of its neighbour, except when this one is far behind it (*i.e.*, the small regions of the opposite colour for  $d_{ij}^{\text{FB}} < -2\text{BL}$ ). In *D. rerio*, the colour of each vertical half-plane of Fig. 3C is almost uniform, except for some regions close to the focal fish ( $d_{ij}^{\text{LR}} \approx \pm 1\text{BL}$ ) at  $\phi_{ij} \approx \pm 90^\circ$ , and some distant regions located at  $d_{ij}^{\text{LR}} \approx \pm 3\text{BL}$  and  $\phi_{ij} \approx \pm 135^\circ$ . This means that the focal fish turns almost always towards its neighbour, and almost independently from its relative heading, except when both fish are close and perpendicular to each other, and when they are very far and almost anti-aligned. In Fig. 3D, one can see that the fish turns to adopt the same direction as that of its neighbour when this one is close and in front of it (*i.e.*, the large green and orange homogeneous regions in the centre of the figure). But the fish tends to turn towards the opposite direction when its neighbour is far ahead (small regions where  $d_{ij}^{\text{FB}} > 2\text{BL}$ ) or behind it and not very close to it (regions where  $d_{ij}^{\text{FB}} < -1\text{BL}$ ). This is a different and more complex behaviour than the one observed in *H. rhodostomus*, where heading changes depend on  $\phi_{ij}$  but not on  $d_{ij}^{\text{LR}}$ . In *D. rerio*, we observe the opposite, and the dependence on the FB distance is more complex than in *H. rhodostomus*. However, neither  $d_{ij}^{\text{LR}}$ , nor  $d_{ij}^{\text{FB}}$ , separately, can determine the distance at which the neighbour is (*e.g.*, a fish  $j$  located at  $1\text{BL}$  to the left can be at  $0.5$  or  $2\text{BL}$  to the front). Therefore, this kind of representation hides the effect of the absolute intensity of the interactions as a function of the distance between fish, and moreover does not allow to disentangle the contribution of intermediate effects such as attraction and alignment.

## 2.2 Limitations of force maps

The above descriptions show that the use of force maps to characterize interactions between individuals raises several problems, especially when more than two state

variables must be taken into account to describe the behaviour of fish. Other issues are not exclusive of force maps. For example, data can be scarce; not only because they are difficult to collect, but also because the phenomenon under observation rarely produce data of a given kind. This is what happens for instance when we consider the repulsive interactions: as individuals repel each other, there are very few cases in which the distance between them is small, so that repulsive interactions are very difficult to describe when the distance between fish is small. This particular issue is common to other methods, including ours. More importantly, data are rarely homogeneously distributed, so that different regions of the map can result from averaging a very disparate number of data, meaning that similar colour intensities do not have the same relevance. For instance, Fig. 4 shows that the relative position of the neighbour of fish  $i$  is not homogeneously distributed on the  $(d, \psi)$ -plane, and is differently distributed in both panels, so that information from, *e.g.*, the frontal region in Panel A is more relevant to describe the behaviour of *H. rhodostomus*, than the information from the same region in Panel B to describe the behaviour of *D. rerio*.

Thus, the relative simplicity to visualise and interpret data with force maps comes at the cost of important limitations and oversimplifications. Four of the most critical of these limitations are the following:

- i)* The reduced number of state variables involved in individual behaviour that force maps can handle is limited to 2, at most 3. This limitation leads to the use of projections and average values that can hide crucial features of the phenomenon under study. Moreover, when data are not homogeneously distributed, the same limitation can even produce wrong or inaccurate observations or conclusions.
- ii)* The difficulty of identifying and disentangling intermediate contributions such as attraction and alignment to behavioural patterns;
- iii)* The difficulty of finding simple analytical expressions of the interaction functions in order to implement them in a mathematical model;
- iv)* The difficulty of distinguishing the effects of variables (*e.g.*, the distance between fish  $d_{ij}$ ) from the effect of parameters (*e.g.*, the body length of fish).

Let us describe these limitations in more detail.

(*i*) When a function depends on more than two variables, force maps are mere projections of the function on a 3D surface, where the value of the function has been averaged with respect to one or more variables. However, averaging raises two problems. First, the function can be odd with respect to the variable used to calculate the average, as it is the case of the

heading change  $\delta\phi$  with respect to the angle of perception  $\psi_{ij}$  and the relative heading  $\phi_{ij}$  in *H. rhodostomus*, where  $\delta\phi(d_{ij}, \psi_{ij}, \phi_{ij}) = -\delta\phi(d_{ij}, -\psi_{ij}, \phi_{ij}) = -\delta\phi(d_{ij}, \psi_{ij}, -\phi_{ij})$  [12]. Then, for instance, huge variations of  $\delta\phi$  with respect to  $\psi_{ij}$ , but of different sign, can cancel each other and yield a small average value, as if  $\delta\phi$  was almost independent of  $\psi_{ij}$ . In other words, a crucial feature of the behaviour can be hidden by the averaging process. Second, averaging by simply adding the values and dividing by the number of values implicitly assumes that the probability of occurrence is the same for all the possible states.<sup>1</sup> This would mean for instance that the probability for a fish  $j$  of being at the state  $d_{ij} = 1 \text{ BL}$ ,  $\psi_{ij} = 90^\circ$  and  $\phi_{ij} = 10^\circ$  with respect to a focal fish  $i$ , is the same than the probability of being at  $(d_{ij}, \psi_{ij}, \phi_{ij}) = (2, 0, 180)$ , which is clearly not the case, at least in *H. rhodostomus* [12].

Thus, even if averaging over some state variables to obtain a 2D force map can provide qualitative information, it can also hide crucial effects and can even produce a completely misleading result due to the strong correlations between the states variables along the actual trajectories. For instance, averaging the interaction between 2 fish (and neglecting the effect of the wall) over their distance  $d$ , to only keep the dependence on the two angular variables  $\psi$  (viewing angle) and  $\phi$  (heading difference) in a 2D representation, involves the unknown correlation function  $C(d, \psi, \phi)$  between  $d$ ,  $\psi$ , and  $\phi$  along the experimental trajectories. Note that even if the true interactions were in fact separable into a product of one-variable interaction functions,  $\delta\phi = f(d)g(\psi)h(\phi)$ , the 2D projection would then be  $\delta\phi(\psi, \phi) = g(\psi)h(\phi) \times \int_0^\infty C(r, \psi, \phi)f(r) dr$ . The integral term (averaging over the distance between the two fish) leads to an unknown and non trivial function of  $\psi$ , and  $\phi$ , which is in general not separable unless the correlation function is itself separable... And it is *not*: if 2 fish are close, they are likely to be aligned ( $\phi \approx 0$ ), whereas if they are far, their headings are much less correlated ( $\phi$  nearly uniformly distributed). This simple example demonstrates that force map projections not only *do not recover the true interaction* ( $g(\psi)h(\phi)$  in this case), but can artificially produce a non-separable interaction, even if the actual interaction takes a product form!

As a consequence, the small number of dimensions that can be represented by force maps constitutes a crucial limitation for an accurate description of behaviour, especially if more variables are taken into account, as, *e.g.*, the relative speed  $v_{ij}$ , or the interaction of fish with an obstacle (such as the wall of a tank), which would require two more variables  $r_{w,i}$  and  $\theta_{w,i}$ , the distance and angle to the wall, respectively. Even for the depicted intermediate values, the precise contributions of each variable on the heading change are still entangled. For instance, how does  $\phi_i$  vary when  $j$  points to the left

<sup>1</sup>The average function of  $a(x, y)$  with respect to the variable  $x$  is given by  $\langle a \rangle_x(y) = \int_x p(x, y)a(x, y)dx$ , where  $p(x, y)$  is the probability of occurrence of the state  $(x, y)$ . However, force maps calculate the  $\langle a \rangle_x(y)$  as the mean of the values of  $a(x, y)$  over all the values of  $x$ , *i.e.*,  $\langle a \rangle_x(y) = \sum_x a(x, y) / \sum_x 1$ , as if  $p(x, y) = 1$  for all  $(x, y)$ , because knowing  $p(x, y)$  is part of the problem.

and is located at the right side of  $i$ , and how does this change vary with the distance  $d_{ij}$ ? The contribution of each variable cannot be disentangled from those of the other variables.

Moreover, force maps of 3D-functions like  $\delta\phi$  require huge amounts of data. Using 30 bins per variable with an average of 100 points per bin would require  $2.7 \times 10^6$  data points, which is very far from being the amount of data collected in most of the experiments on collective motion. We anticipate here that, in order to get the same level of precision, our procedure described below reduces the required number of data up to 150 times.

(ii) Another limitation of force maps is that intermediate contributions are difficult or often impossible to identify. A function  $f$  can depend on the state variables  $x$  and  $y$  through two intermediate functions  $a$  and  $b$  such that  $f(x, y) = f(a(x, y), b(x, y))$ . This is precisely the case of the heading change  $\delta\phi$ , which depends on the state variables  $d_{ij}$ ,  $\psi_{ij}$  and  $\phi_{ij}$  through the combination of at least two intermediate contributions, attraction and alignment, themselves depending on the three state variables [12]. Moreover, attraction and alignment can have opposite contributions, so that their combined effect can be cancelled as if the fish were isolated from each other, while both forces are in action. Force maps are not able to identify which part of the heading change is due to the attraction or the alignment, that is, the functions  $a(x, y)$  and  $b(x, y)$  cannot be directly extracted from force maps. Even if representations in 7D were possible, these “maps” would not allow to identify intermediate contributions.

(iii) Extracting analytical expressions from a colour map is difficult unless the relation between variables is very simple. However, it is essential to build mathematical models to understand how the combination of social interactions gives rise to the observed behaviour [14, 21]. The simulations of these models will then be used to make predictions in other experimental situations, to draw phase portraits, etc. Simple piecewise linear functions interpolating the data are not suitable because they lack the physical or biological meaning and do not help to build explicit and concise mathematical models. Fig. 2 and 3 show the force maps of  $\delta\phi$  for two different species of fish. Significant differences appear between maps, *e.g.*, in Panels A, the left-right symmetry of the heading change with respect to the position of the neighbour in *H. rhodostomus*, while in *D. rerio* the symmetry is with respect to the position of the focal fish; in Panels C, the half-planes of homogeneous colour are horizontal in *H. rhodostomus*, but vertical in *D. rerio*. Will such a difference be observed if larger groups are considered? Force maps cannot be extrapolated from one experimental situation to another.

(iv) Finally, it is difficult to distinguish the effect of a variable from the effect of a parameter in a force map, *e.g.*, the effect of the instantaneous distance between fish at each instant of time, from the effect of the fixed body length of individuals. Are the differences observed in the previous maps of each species due to the experi-

mental conditions (*e.g.*, the radius of the arena, which is a variable of the experimental setup), or to the physical characteristics of the species (*e.g.*, the body length, which is a fixed parameter of the species)?

The method we present below does not have these limitations, first, because it can handle a large number of state variables (*i.e.*, of dimensions), and second, because the analytic expressions it provides are precisely the optimal way to disentangle the interaction functions at play and to describe the role of each state variable and each parameter of the system. These analytic expressions can then be exploited to build an explicit and yet concise model, whose agreement with experiment can be tested, and whose predictions can be further investigated experimentally.

### 3 Method to extract and model social interactions from behavioural data

Section 2 shows that force maps are representations of one quantity (action force, acceleration, heading variation, etc.) as a function of pairs of other quantities (relative position, velocity or orientation, angle of perception of other individuals, etc.). These quantities only make sense in a framework of the physical world described by a mathematical model which, even if it is often not mentioned explicitly in studies [18, 19, 22, 23] (but see also [13, 24, 25]), is usually based on equations of motion built in analogy with Newtonian mechanics.

The method consists essentially in defining this framework and deriving the mathematical model describing the relation between the quantities used to quantify the behavior of an organism and its interactions with other organisms or physical objects that are present in its environment.

Of course, our procedure requires the measurements (*i.e.*, the distribution) of the heading changes  $\delta\phi$ , and also to calculate the average in each box of the discretised grid. However, our procedure does not require at any moment to represent heading changes as a function of two or more variables, *i.e.*, using force maps.

#### 3.1 Outline of the method

Fig. 5 provides a general overview of the method used to design the mathematical model and to extract the social interactions functions. The method consists of five consecutive steps, starting from experimental observations and ending with numerical simulations of a model reproducing these observations. These five steps are repeated in a cycle as shown in Fig. 5, as the model predictions allow to perform better targeted experiments, which in turn allow to refine the model: the experiments feed the model, and in turn the model helps the design of experiments. Fig. 6 shows the same five steps in a detailed flow chart that includes the test of the model.

### 3.1.1 Data collection and preparation

The method starts by carrying out experiments, motivated by a fundamental question, or the results of the previous cycle of an ongoing modelling process. Then, data are collected and prepared in the form of individual trajectories  $\vec{u}_i(t)$  resulting from a tracking procedure conveniently purged from identification errors (step 1 in Figs. 5 and 6). The velocity vector  $\vec{v}_i$  of an individual  $i$  is calculated from successive positions with finite differences, and its heading  $\phi_i$  is extracted directly from the tracking data, or, alternatively, taken as the orientation of the velocity vector.

### 3.1.2 Identification and analysis of state variables and observables

The second step consists in identifying and analysing the state variables and the relevant observables that can be used to describe the phenomenon under study (step 2 in Figs. 5 and 6).

In the case considered here, the individual state of a fish is characterized by its position and velocity vectors, or, equivalently, by its speed and its distance and orientation to the wall,  $v$ ,  $r_w$ , and  $\theta_w$ , respectively. The state of a fish  $i$  with respect to another fish  $j$  is characterized by the distance between them  $d_{ij}$ , the relative speed  $v_{ij}$ , the viewing angle with which  $j$  is perceived by  $i$ ,  $\psi_{ij}$  ( $\neq \psi_{ji}$ ), and their relative heading  $\phi_{ij} = \phi_j - \phi_i$ . The heading angle change  $\delta\phi_S$  resulting from social interactions is thus a function of four variables.

Then, an analysis is performed on a series of observables that can be used to describe both the individual and collective behaviour of fish when swimming in pairs. These observables are the probability density function (PDF) of the spatial distribution of individuals, their velocity, their distance and orientation to obstacles, their heading variation, the distance between individuals (cohesion), their relative positions, and their relative heading (polarization). Fig. 7 shows the PDF of  $d_{ij}$ ,  $\psi_{ij}$  and  $\phi_i$  corresponding to the species of fish studied here. The analysis of these observables relies exclusively on the experimental data and is independent of a model. Eventually, force maps can be drawn.

### 3.1.3 Modelling hypotheses

The next step consists in defining a class of models that can provide the more suitable description of the observed phenomenon (step 3 in Figs. 5 and 6). This is by far the crucial part of the method.

This process has been described in detail in [12] in the case of *H. rhodostomus*; here we simply summarise it, as it is identical for *D. rerio*. We first consider that fish move straight during short time intervals of different duration. These time intervals are separated by instantaneous ‘‘kicks’’ during which fish adjust their direction and make an abrupt acceleration, after which the speed decays exponentially as the fish glides between two kicks. These assumptions come from the analysis

of the trajectories and the variation of the velocity [12]. We also assume that the effects of social interactions occur exclusively when kicks are performed, that is, at the decision times when fish adjust their heading. Then, we provide an explicit equation for the heading variation of the focal fish  $\delta\phi$  when the fish performs a kick, in terms of the quantities that can potentially have an effect on  $\delta\phi$ . The selection of the duration and length of the gliding phase and the motion of the fish, with quasi-exponential decay of the velocity between two kicks, are given by simple analytical probability distributions fairly reproducing the experimental ones, as described in [12].

The equations for the time-evolution of the vector position  $\vec{u}(t)$  and the heading  $\phi(t)$  of a fish are

$$\begin{aligned}\vec{u}(t + dt) &= \vec{u}(t) + l(t) \vec{e}(\phi(t + dt)), \\ \phi(t + dt) &= \phi(t) + \delta\phi(t),\end{aligned}$$

where  $l(t)$  is the length of the kick, performed in the direction of the angle  $\phi(t + dt)$ , given by the unitary vector  $\vec{e}(\phi(t + dt))$ . The choice of the kick length is described in detail in [12]. Here we focus on the equation for the heading variation  $\delta\phi$ , that, we hypothesize, accounts for the effects of the social interactions between fish, and reads

$$\delta\phi(t) = \delta\phi_R(t) + \delta\phi_w(t) + \delta\phi_S(t), \quad (1)$$

where  $\delta\phi_R$  is a random angle change accounting for the spontaneous decisions of the fish,  $\delta\phi_w$  is due to the repulsion of the wall when the fish is close to a tank wall, and  $\delta\phi_S$  is due to the social interactions with other fish, usually attraction, alignment, or a combination of both, so that

$$\delta\phi_S(t) = \delta\phi_{Att}(t) + \delta\phi_{Ali}(t). \quad (2)$$

Eqs. (1)-(2) are based on the hypotheses that  $\delta\phi_R$ ,  $\delta\phi_w$ ,  $\delta\phi_{Att}$ , and  $\delta\phi_{Ali}$  are the main contributions to the heading variation of a fish, and that these contributions are combined linearly (additive hypothesis).

### 3.1.4 Extraction of interactions functions from trajectory data

Step 4, which is the one we wish to emphasize in this article, consists in extracting, from the experimental data, the interaction functions determining the contributions of a neighbouring fish and of the obstacles in the environment to the instantaneous heading variation of a fish. It is a relatively long step that involves several substeps, for which an overview is presented in Fig. 8.

Discretising the 6 state variables in  $\Omega = Q \times P \times I \times J \times K \times M$  boxes, and calculating the mean value of  $\delta\phi$  in each box from the experimental data, Eq. (1) can be rewritten as a system of  $\Omega$  equations and  $\Omega$  unknowns,

$$\delta\phi_w(r_q, \theta_p) + \delta\phi_S(d_i, \psi_j, \phi_k, v_m) = \delta\phi_{qpijkm}, \quad (3)$$

where the noise term  $\delta\phi_R$  vanishes as it is assumed to have zero mean in each box. We used the convention that unknowns are written in the left hand side of the

633 equation and the known values in the right hand side.  
 634 Even if one could build seven-dimensional representations,  
 635 the contributions of each variable would still be  
 636 entangled, and intermediate functions impossible to detect  
 637 in the corresponding behavioural maps of social  
 638 interactions.

The key hypothesis to the procedure consists in assuming that the contribution of each state variable to each kind of interaction *can be separated in a product form*, which is the case for physical particles [12]:

$$\delta\phi_w(r_w, \theta_w) = \bar{f}(r_w) \bar{g}(\theta_w), \quad (4)$$

$$\delta\phi_{Att}(d, \psi, \phi, v) = f(d) g(\psi) h(\phi) l(v), \quad (5)$$

$$\delta\phi_{Ali}(d, \psi, \phi, v) = \hat{f}(d) \hat{g}(\psi) \hat{h}(\phi) \hat{l}(v). \quad (6)$$

639 Note that functions with a hat are in general different  
 640 from the functions without hat, (*i.e.*,  $f(d) \neq \hat{f}(d)$ , etc.).  
 641 With this additional hypothesis, the procedure provides  
 642 analytical expressions of the interaction functions of attraction  
 643 and alignment, and of the contribution of each of them to the heading  
 644 variation. In the case where the contribution of one of the interactions  
 645 is non-existent, for instance, if there is no explicit alignment,  
 646 then the procedure will detect it.

648 For the sake of simplicity, we consider here the case  
 649 where fish are far enough from the tank wall so that the contribution  
 650 of  $\delta\phi_w$  to the heading variation can be neglected with respect to the  
 651 effects of social interactions. The procedure can easily be extended to  
 652 extract both the social interactions and the interaction with the wall,  
 653 but to the cost of more complicated notations. We refer the reader to  
 654 the work of Calovi *et al.* [12], in which the disentangling of the  
 655 combined effects of a tank wall and social interactions between fish  
 656 is described in detail.

658 Hence, we are left with 4 state variables, so that the heading change  
 659  $\delta\phi(t)$  is averaged in 4-dimensional  $ijkm$ -boxes. We introduce the  
 660 notation  $f(d_i) = f_i$ ,  $g(\psi_j) = g_j$ , ...,  $\hat{l}(v_m) = \hat{l}_m$ , so that, after  
 661 removing the interaction with the wall, the system (3) becomes

$$f_i g_j h_k l_m + \hat{f}_i \hat{g}_j \hat{h}_k \hat{l}_m = \delta\phi_{ijkm}, \quad (7)$$

663 which is a system of  $I \times J \times K \times M$  equations with  
 664  $D = 2(I + J + K + M)$  unknowns, which is much smaller than the  
 665 number of boxes, and in practice, to the usual number of data available.

667 The goal is now to solve this system of equations. This will provide  
 668 the discrete values of the interaction functions (the unknowns  $f_i$ ,  $g_j$ ,  
 669 ...,  $\hat{l}_m$ ) in function of the known values  $\delta\phi_{ijkm}$  (calculated from the  
 670 experimental data as the mean heading change in each box). Systems  
 671 with more equations than unknowns are called *overdetermined systems*  
 672 and rarely have a solution. One way of overcoming this problem  
 673 consists in reducing the overdetermined system to a solvable one by  
 674 minimizing the error  $\Delta$  with which the overdetermined system is  
 675 satisfied by a candidate solution that is updated iteratively. This  
 676 corresponds to the step 4.1 in the flowchart of Fig. 6, which is  
 677 described in detail in Box 1 and in the

680 next section 3.2 for a simple 2D-case where the equation is  
 681  $f_i g_j = a_{ij}$ . Fig. 9 shows how the error function is built in this  
 682 simple case.

The error function is a function  $\Delta(\vec{X})$  of  $D = 2(I + J + K + M)$   
 variables which returns a real value calculated as follows:

$$\Delta(\vec{X}) = \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K \sum_{m=1}^M \epsilon_{ijkm} (\delta\phi_{S,ijkm} - \delta\phi_{ijkm})^2, \quad (8)$$

683 where  $\vec{X}$  is the vector of dimension  $D$

$$\vec{X} = (f_1, \dots, f_I, g_1, \dots, g_J, h_1, \dots, h_K, l_1, \dots, l_M, \hat{f}_1, \dots, \hat{f}_I, \hat{g}_1, \dots, \hat{g}_J, \hat{h}_1, \dots, \hat{h}_K, \hat{l}_1, \dots, \hat{l}_M),$$

684 and where  $\delta\phi_{S,ijkm} = f_i g_j h_k l_m + \hat{f}_i \hat{g}_j \hat{h}_k \hat{l}_m$  is the discretisation  
 685 of the social force in each  $ijkm$ -box (the unknowns),  $\delta\phi_{ijkm}$  is the mean heading  
 686 change in the  $ijkm$ -box (calculated from experimental data), and  
 687  $\epsilon_{ijkm}$  is the number of data in the  $ijkm$ -box. By weighting the regions  
 688 with more data, the factor  $\epsilon_{ijkm}$  allows to preserve the structure of the  
 689 dataset and the correlations between variables resulting from the dynamics  
 690 (see also note 1 in Sec. 2). For a given  $\vec{X}$ , the local error in the  
 691  $ijkm$ -box is defined as the difference between  $\vec{X}$  and the experimental data,  
 692  $\delta\phi_{S,ijkm} - \delta\phi_{ijkm}$ . Then, the local error is squared and weighted by the  
 693 number of data in the box,  $\epsilon_{ijkm}$ , and summed up over all the  $I \times J \times K \times M$   
 694 boxes. The result is a positive real value  $\Delta(\vec{X})$  that must be as small as  
 695 possible, so that an optimal  $\vec{X}$  must be found.

697 The minima of  $\Delta(\vec{X})$  are obtained by finding the zeros of the gradient  
 698 of  $\Delta$  (see the inset in Box 1). Finding the zeros of a function (or *root finding*)  
 699 can be carried out by different methods such as descent methods or other  
 700 iterative methods. Here, we use an iterative method based on the fact that  
 701 the equation that each component must verify is linear in this component,  
 702 so that each component can be written explicitly in terms of the other  
 703 components. This relation is found as follows.

704 For the gradient to be zero, we must have

$$\frac{\partial \Delta}{\partial f_1} = 0, \quad \frac{\partial \Delta}{\partial f_2} = 0, \quad \dots, \quad \frac{\partial \Delta}{\partial f_I} = 0, \quad \dots, \quad \frac{\partial \Delta}{\partial l_M} = 0.$$

705 For the component  $f_i$ , the partial derivative of  $\Delta$  is

$$\frac{\partial \Delta}{\partial f_i} = 2 \sum_{j=1}^J \sum_{k=1}^K \sum_{m=1}^M \epsilon_{ijkm} g_j h_k l_m (\delta\phi_{S,ijkm} - \delta\phi_{ijkm}),$$

so the condition  $\partial\Delta/\partial f_i = 0$  is equivalent to

$$f_i = \frac{\sum_{j=1}^J \sum_{k=1}^K \sum_{m=1}^M \epsilon_{ijkm} g_j h_k l_m (\hat{f}_i \hat{g}_j \hat{h}_k \hat{l}_m - \delta\phi_{ijkm})}{\sum_{j=1}^J \sum_{k=1}^K \sum_{m=1}^M \epsilon_{ijkm} (g_j h_k l_m)^2}. \quad (9)$$

712 The corresponding equations for the other components  
 713 are derived in the same way and all have the same struc-  
 714 ture where each component is given by an explicit com-  
 715 bination of the other components and the known values  
 716  $\delta\phi_{ijkm}$  and  $\epsilon_{ijkm}$ . For instance, the equation of  $g_j$   
 717 is obtained by replacing “ $g_j$ ” by “ $f_i$ ”, “ $j = 1$ ” by “ $i = 1$ ”,  
 718 and “ $J$ ” by “ $I$ ”, and the equation for a function with  
 719 a hat (e.g., “ $\hat{f}_i$ ”) is obtained by replacing the functions  
 720 that have a hat by the same functions without the hat,  
 721 and *vice versa* (i.e., “ $g_j$ ” is replaced by “ $\hat{g}_j$ ”, “ $\hat{g}_j$ ” is  
 722 replaced by “ $g_j$ ”, and so on...).

723 The resulting set of explicit equations for each com-  
 724 ponent can be written in the following form,

$$\begin{aligned}
 f_1 &= R_1(g_1, \dots, g_J, h_1, \dots, \hat{h}_K, \hat{l}_1, \dots, \hat{l}_M), \\
 f_2 &= R_2(g_1, \dots, g_J, h_1, \dots, \hat{h}_K, \hat{l}_1, \dots, \hat{l}_M), \\
 &\vdots = \vdots \\
 f_I &= R_I(g_1, \dots, g_J, h_1, \dots, \hat{h}_K, \hat{l}_1, \dots, \hat{l}_M), \\
 g_1 &= R_{I+1}(f_1, \dots, f_I, h_1, \dots, \hat{h}_K, \hat{l}_1, \dots, \hat{l}_M), \\
 &\vdots = \vdots \\
 \hat{g}_J &= R_{I+J}(f_1, \dots, f_I, h_1, \dots, \hat{h}_K, \hat{l}_1, \dots, \hat{l}_M), \\
 &\vdots = \vdots \\
 \hat{l}_M &= R_D(f_1, \dots, f_I, g_1, \dots, \hat{g}_J, \hat{h}_1, \dots, \hat{h}_K).
 \end{aligned}$$

725 To find the solution of such a system, it is possible to  
 726 write it in a compact form as  $\vec{X} = \vec{R}(\vec{X})$ , where  $\vec{R}$   
 727 is a function from  $\mathbb{R}^D$  to  $\mathbb{R}^D$ , and consider the problem  
 728 as finding a fixed point  $\vec{X}_0$  of the function  $\vec{R}$ . This  
 729 is done by means of an iterative method, as explained  
 730 in detail in Box 2, which also shows how this method  
 731 works in the 1D-case. Again, note that the number  
 732  $D = 2(I + J + K + M)$  of fitting parameters is much  
 733 smaller than the number of boxes  $I \times J \times K \times M$ .

734 In the end, the procedure provides the values of each  
 735 interaction function at the points representing each box  
 736 minimizing the error functions. Depending on the num-  
 737 ber of boxes, a satisfactory analytical form of each one-  
 738 variable function can be obtained, based on physical  
 739 principles and specific observations. This part corre-  
 740 sponds to the step 4.3 in Fig. 6 and depends on the spe-  
 741 cific phenomenon under study; it is detailed in Sec. 4  
 742 for the case of fish swimming in pairs. For example,  
 743 the function of the angle of perception  $g(\psi)$  must be  
 744 odd, i.e.,  $g(-\psi) = -g(\psi)$ , because the attractive effect  
 745 of a neighbour on the heading change of a focal fish  
 746 has the same intensity wherever the neighbour is lo-  
 747 cated at the right or at the left of the focal fish, but has  
 748 opposite directions, and hence opposite signs, in each  
 749 case: if the neighbour is at the right (resp. left) side,  
 750 the focal fish would turn right (resp. left) to approach  
 751 the neighbour. Physical properties of the interactions  
 752 at play, such as the exponential decay of some interac-  
 753 tions, must be taken into account and guide the choice  
 754 of the final analytical expressions.

### 755 3.1.5 Numerical simulations of the model

756 The last step consists in performing numerical simula-  
 757 tions of the model with the double purpose of 1) veri-  
 758 fying that the simulation results are in good agreement  
 759 with the experimental results, and 2) making new pre-  
 760 dictions that can be ultimately confirmed by new exper-  
 761 iments. If one of these two points is not satisfactorily  
 762 verified, then it is necessary to go back to a previous  
 763 step: step 3 to revise or reformulate the model, step 2  
 764 to use alternative state variables or observables, or even  
 765 step 1 to carry out new experiments or measures of the  
 766 data (Fig. 6).

## 767 3.2 Application to a simple case study

768 In order to illustrate how the extraction procedure  
 769 (step 4 of our methodology in Fig. 6) can be used, we  
 770 have produced artificial data for a simple case in which  
 771 the interaction functions are known and in which we  
 772 controlled the level of noise and the distribution of data.  
 773 This simple case also allows us to illustrate the efficiency  
 774 of the procedure and the accuracy that can be obtained  
 775 according to the quality of the data.

776 Figs. 10AB show the colour map of a real function  
 777 of two variables  $a(x, y)$  going from  $[0, L] \times [-\pi, \pi]$  to  $\mathbb{R}$ ,  
 778 for two levels of noise:

$$779 \quad a(x, y) = \sqrt{2}e^{-(x/x_0)^2} \sin y, \quad (10)$$

780 with  $x_0 = 0.1$  and  $L = 2$ . The map is built as follows.  
 781 We first discretise the 2D-space  $(x, y)$  in  $\Omega = I \times J$   
 782 rectangular cells  $[\hat{x}_i, \hat{x}_{i+1}] \times [\hat{y}_j, \hat{y}_{j+1}]$ , where

$$\begin{aligned}
 \hat{x}_i &= (i - 1) \frac{L}{I}, \quad i = 1, \dots, I + 1, \\
 \hat{y}_j &= -\pi + (j - 1) \frac{2\pi}{J}, \quad j = 1, \dots, J + 1.
 \end{aligned}$$

783 Then, we evaluate the function  $a(x, y)$  on 100000 points  
 784 randomly selected, and place each point on the corres-  
 785 ponding cell  $ij$ . After that, we count the number of  
 786 points in each cell,  $\epsilon_{ij}$ , and we assign to  $a_{ij}$  the aver-  
 787 age of the values of the function for all the points that  
 788 are in the cell  $ij$ . To simulate the effect of the noise,  
 789 which is always present in real data sets, for each point  
 790 found in the cell  $ij$ , we add a small noise of zero mean  
 791 and standard deviation = 0.35 to the value of  $a_{ij}$ . This  
 792 value corresponds to the noise intensity observed in the  
 793 experiments with pairs of real fish [12]. The resulting  
 794 value is then considered as the *measured* value of the  
 795 function  $a(x, y)$  in the corresponding cell, denoted by  
 796  $(x_i, y_j)$ , and usually defined as the middle point of the  
 797 cell:  $x_i = (\hat{x}_{i+1} - \hat{x}_i)/2$ ,  $y_j = (\hat{y}_{j+1} - \hat{y}_j)/2$ .

798 We now apply the procedure described in Sec. 3.1.4,  
 799 which provides analytical expressions of the interaction  
 functions of  $x$  and  $y$  that give rise to this colour map.

The key hypothesis is that there exist two decoupled  
 functions  $f(x)$  and  $g(y)$  such that  $a(x, y) = f(x)g(y)$ .  
 Evaluating these functions in each  $ij$ -cell, this means

that

$$f_i g_j = a_{ij}, \quad i = 1, \dots, I, \quad j = 1, \dots, J. \quad (11)$$

The values of  $f_i$  and  $g_j$  are unknown for all  $i$  and all  $j$ . The values of  $a_{ij}$  are known for all  $ij$  because it is the mean value of the data found in the  $ij$ -cell. These  $I \times J$  equations and  $I + J$  unknowns constitute precisely the overdetermined system (7). Following the procedure to reduce this overdetermined system, we write the error function  $\Delta(\vec{X})$ , for  $\vec{X} = (f_1, \dots, f_I, g_1, \dots, g_J)$ ,

$$\Delta(\vec{X}) = \sum_{i=1}^I \sum_{j=1}^J \epsilon_{ij} (f_i g_j - a_{ij})^2, \quad (12)$$

800 which has the same form as the one shown in Eq. (27)  
801 of Box 1. Fig. 9 shows how  $\Delta(\vec{X})$  is calculated as the  
802 sum of all the squared local errors in each box, denoted  
803 by the vertical green line between the yellow and green  
804 balls corresponding to  $a_{ij}$  and  $f_i g_j$  respectively.

To minimize the error function, we look for the zeros of its gradient. The partial derivatives are

$$\frac{\partial \Delta}{\partial f_i} = 2 \sum_{j=1}^J \epsilon_{ij} g_j (f_i g_j - a_{ij}), \quad (13)$$

$$\frac{\partial \Delta}{\partial g_j} = 2 \sum_{i=1}^I \epsilon_{ij} f_i (f_i g_j - a_{ij}), \quad (14)$$

and they are equal to zero when

$$f_i = \sum_{j=1}^J \epsilon_{ij} g_j a_{ij} / \sum_{j=1}^J \epsilon_{ij} g_j^2, \quad (15)$$

$$g_j = \sum_{i=1}^I \epsilon_{ij} f_i a_{ij} / \sum_{i=1}^I \epsilon_{ij} f_i^2. \quad (16)$$

805 Note that each component is explicitly given by a combination  
806 of the other components. This property always holds, independently  
807 of the number of interaction functions. This is due to the hypothesis  
808 of separability of variables and to the square norm used in the error  
809 function. Moreover, in this particularly simple case, we observe  
810 that the “ $f_i$ ’s” are given by the “ $g_j$ ’s”, and vice versa (recall  
811 that the values  $a_{ij}$  and  $\epsilon_{ij}$  are known for all  $ij$ ). Note finally  
812 that the minus sign that appears in Eq. (9) does not appear here.

815 Thus, starting from an initial guess of the solution,  
816  $\vec{X}^0 = (f_1^0, f_2^0, \dots, f_I^0, g_1^0, g_2^0, \dots, g_J^0)$ , it is possible to  
817 obtain the values of the  $f_i$ ’s and the  $g_j$ ’s after one iteration,  
818 that is,  $\vec{X}^1 = (f_1^1, f_2^1, \dots, f_I^1, g_1^1, g_2^1, \dots, g_J^1)$ , repeating  
819 this process iteratively,  $\vec{X}^1 \rightarrow \vec{X}^2 \rightarrow \vec{X}^3 \rightarrow \dots$  until  
820 convergence, *i.e.*,  $\vec{X}^{n+1} \approx \vec{X}^n$ .

When the jumps between iterations are excessively abrupt, for example if the initial guess is far from a solution, it is convenient to smooth or *relax* the iterative process by averaging the new values with those of the previous step weighted with a larger coefficient:

$$f_i^{n+1} = \lambda f_i^n + (1 - \lambda) \tilde{f}_i^n, \quad (17)$$

$$g_j^{n+1} = \lambda g_j^n + (1 - \lambda) \tilde{g}_j^n. \quad (18)$$

821 Here  $\tilde{f}_i^n$  and  $\tilde{g}_j^n$  are calculated from the values of  $\vec{X}^n$   
822 in the previous step with the formulas (15)-(16), and  
823  $\lambda \in [0, 1]$  is chosen with the compromise that the iterations  
824 progress smoothly but in a reasonable computational time. See also Box 2.

826 Fig. 11 shows the successive values that each component  $f_i$ ,  $g_j$   
827 adopt during the first 120 steps of the iteration process, starting  
828 from an initial condition where  $f_i^0 = g_j^0 = 1$  for all  $i$  and  $j$ . We  
829 used  $I = 11$ ,  $J = 17$ , a value of  $\lambda = 0.975$  quite close to 1 in  
830 order to illustrate the convergence process in detail, and a final  
831 tolerance  $|\Delta(\vec{X}^{n+1}) - \Delta(\vec{X}^n)| < \varepsilon = 1.87 \times 10^{-7}$  (using  
832  $\varepsilon = I \times J \times 10^{-9}$ ). Typical values of  $\lambda$  can be much  
833 smaller ( $\lambda = 0.75$ ), depending on how close the initial  
834 guess is from the solution.

836 The final state to which the iterative method has converged, that  
837 is, the points  $\{f_i\}_{i=1}^I$  and  $\{g_j\}_{j=1}^J$  that solve the reduced  
838 system (31), is shown in Figs. 10CD. There is an excellent  
839 agreement with the original data, despite the addition of noise,  
840 including in the case of a much larger noise than the one  
841 inferred in the actual data of our fish experiments.

843 Note that if  $f$  and  $g$  are a solution of (31), then the functions  
844  $(1/\alpha)f$  and  $\alpha g$ , where  $\alpha$  is a real number, are also a solution  
845 of (31), since the product of the two functions remains invariant.  
846 Hence, we need to impose an additional condition in order to  
847 account for this under-determination and to generally allow for a  
848 proper comparison of reconstructed interaction functions. Following  
849 [12], we chose to normalize all angular functions so that their  
850 squared average is equal to 1. In particular,  $g$  is normalized  
851 such that  $(1/2\pi) \int_{-\pi}^{\pi} g(y)^2 dy = 1$ , a normalization applied  
852 in Figs. 10CD.

854 The last step consists in finding simple analytical expressions  
855 that interpolate the discrete values of the reconstructed interaction  
856 functions, in order to implement them in an explicit mathematical  
857 model. There is an infinite number of combinations, so that one  
858 must be guided by key physical features of the phenomenon under  
859 study that are well established, properties such as symmetries,  
860 analogies with other physical systems. For example, in the case  
861 of angular functions, the parity is often easily identifiable from  
862 the data or can be asserted from general principles (mirror  
863 symmetry, left-/right symmetry...), so that few Fourier modes  
864 can be sufficient to interpolate the angular functions from the  
865 data that result from the reconstruction procedure.

### 868 3.3 On the hypothesis of separation of variables of interaction functions

870 The central assumption of the extraction procedure is the  
871 separation of variables made in (4)-(6). Without this hypothesis,  
872 the solution of Eq. (3) far from the wall is  $\delta\phi_S(d_i, \psi_j, \phi_k, v_m) = \delta\phi_{ijk m}$ , and we  
873 are lead back to force maps, with all contributions still entangled  
874 and no analytical expressions of the interaction functions. Variables  
875 separability is thus crucial for our method.

877 However, what happens if the interaction functions do  
878 not satisfy this condition?

Consider a 2D-function  $a(d, \psi)$ , odd in  $\psi$ . Such a function can always be expanded in Fourier series as

$$a(d, \psi) = \sin \psi \sum_{k=0}^{\infty} c_k(d) \cos(k\psi). \quad (19)$$

879 In order to be separable as  $a(d, \psi) = f(d)g(\psi)$ , the func-  
880 tions  $c_k(d)$  in (19) have all to be proportional to some  
881  $f(d)$ , that is,  $c_k(d) = \alpha_k f(d)$ , with  $\alpha_k$  constant for all  $k$ .  
882 Then,  $g(\psi) = \sin \psi \sum_{k=0}^{\infty} \alpha_k \cos(k\psi)$ .

A simple and biologically meaningful example, inspired by the fish interaction functions, would consist in writing  $c_k(d) = \alpha_k \exp[-d^2/(2l_k^2)]$  in (19) and using only the modes  $k = 0$  and 2:

$$a(d, \psi) = \sin \psi \left[ \alpha_0 e^{-d^2/(2l_0^2)} + \alpha_2 e^{-d^2/(2l_2^2)} \cos(2\psi) \right].$$

When  $l_2 = l_0$ , the variables are separable and we have

$$a(d, \psi) = \alpha_0 e^{-d^2/(2l_0^2)} \sin \psi \left[ 1 + \frac{\alpha_2}{\alpha_0} \cos(2\psi) \right], \quad (20)$$

883 which, when  $\alpha_0 = \sqrt{2}$ ,  $l_0 = x_0/\sqrt{2}$ , and  $\alpha_2 = 0$ , cor-  
884 responds to the simple case used in Sec. 3.2, for which  
885 the reconstruction procedure worked quite well.

Let see the result of applying our procedure to a similar function  $b(d, \psi)$  that does not satisfy the separability condition. Using the Fourier expansion (19), we consider two examples  $b_1(d, \psi)$  and  $b_2(d, \psi)$  of the form

$$b_{1,2}(d, \psi) = \sqrt{2} e^{-d^2/(2l_0^2)} \sin \psi + 1.13 e^{-d^2/(2l_2^2)} \sin \psi \cos(2\psi),$$

886 where  $l_2 = 2l_0/3 \approx 0.047$  and  $l_2 = 3l_0/2 \approx 0.106$  respec-  
887 tively, and  $\alpha_2 = 0.8\alpha_0 \approx 1.13$  in both functions. Note  
888 that the first term is precisely the function used in the  
889 simple case of Sec. 3.2, and the presence of the second  
890 term makes the functions  $b_1$  and  $b_2$  not separable.

Figs. 12AB show the colour maps of  $b_1(d, \psi)$  and the solution found by the reconstruction procedure  $f_i g_j$ . The general shape is quite well reproduced, and differences appear only at a relatively small scale. Fig. 12CD shows the reconstructed functions  $f(d)$ ,  $g(\psi)$  for which we have found the following analytical expressions,

$$f(d) = \beta_0 e^{-d^2/[2(\bar{l}_0)^2]}, \quad (21)$$

$$g(\psi) = \beta_1 \sin \psi [1 + \beta_2 \cos(2\psi)], \quad (22)$$

with  $\beta_0 = 0.72$ ,  $\bar{l}_0 = l_0 = 0.071$ ,  $\beta_1 = 1.86$ ,  $\beta_2 = 0.62$ . The resulting interaction function is

$$f(d)g(\psi) = 1.34 e^{-d^2/[2(\bar{l}_0)^2]} \sin \psi [1 + 0.62 \cos(2\psi)],$$

891 where the first term is quite close to the one of  $b_1(d, \psi)$   
892 (with 1.34 instead of 1.41), and the second is not  
893 too far from the corresponding one of  $b_1(d, \psi)$ , with  
894  $1.34 \times 0.62 = 0.83$  instead of 1.13, apart from the dif-  
895 ferent decreasing rate of the exponential.

896 Although the procedure provided an apparently sat-  
897 isfactory result (panels A and B of Fig. 12 are quite  
898 similar), the reconstructed functions fail to reproduce  
899 some features such as the (two) changes of variation of  
900 the intensity along the radial coordinate (first increas-  
901 ing, then decreasing) when the neighbour is at one side  
902 of the focal individual ( $\psi \approx \pm\pi/2$ ), and introduce an  
903 artificial slight decrease in the intensity when the neigh-  
904 bour is far ( $d \approx 3BL$ ) and at one side ( $\psi \approx \pm\pi/2$ ) of  
905 the focal individual (see Fig. 12CD).

A similar qualitative result is obtained in the second case with  $b_2(d, \psi)$ , as shown in Fig. 13. In that case, we found  $\beta_0 = 0.69$ ,  $\bar{l}_0 = 0.078$ ,  $\beta_1 = 1.99$ ,  $\beta_2 = 0.9$ , so the interaction function provided by our procedure is

$$f(d)g(\psi) = 1.37 e^{-d^2/[2(\bar{l}_0)^2]} \sin \psi [1 + 0.9 \cos(2\psi)]$$

906 In both examples, the worst agreement between the re-  
907 constructed interaction and the actual one is obtained  
908 for the angular dependence of the interaction at very  
909 large distance (blue curves in Figs. 12D and 13D),  
910 and for the radial dependence of the interaction at  
911  $\psi = \pm\pi/2$  (red curves in Figs. 12C and 13C). However,  
912 both cases correspond to situations where the interac-  
913 tion is the weakest and the reconstructed interaction  
914 matches perfectly the actual one in situations where the  
915 interaction is most significant (black and red curves in  
916 both panels D; black and blue curves in both panels C).

917 Let us now discuss the general validity of the prod-  
918 uct assumption on the illustrative example of the inter-  
919 action of a fish with a circular wall. In this case, the  
920 heading angle change between two time steps (kicks)  
921 is  $\delta\phi(r_w, \theta_w)$ , where  $r_w$  is the distance of the fish from  
922 the wall, and  $\theta_w$  is the angle between the fish heading  
923 and the normal to the wall. Here, the product hypoth-  
924 esis amounts to assume that  $\delta\phi(r_w, \theta_w) = f(r_w)g(\theta_w)$ .  
925 As explained in [12], the product assumption is ver-  
926 ified for *physical* particles interacting with a wall via a  
927 conservative force (deriving from a potential energy),  
928 with  $g(\theta_w)$  given exactly by  $g(\theta_w) = \sin(\theta_w)$  (projec-  
929 tion of the central force on the normal to the veloc-  
930 ity, i.e., on the angular acceleration). For animals, and  
931 in particular fish, the anisotropic perception of their  
932 environment generally leads to non conservative inter-  
933 actions. The product hypothesis assumes that this  
934 anisotropic perception can be fully encoded in a non-  
935 trivial function  $g(\theta_w)$  generally different from a simple  
936 sinus, while the function remains odd if the left/right  
937 symmetry is preserved. As explained above,  $\delta\phi(r_w, \theta_w)$   
938 for animals takes the general Fourier expansion form  
939 of Eq. (19),  $\delta\phi(r_w, \theta_w) = \sin(\theta_w) \sum_{k=0}^{\infty} c_k(r_w) \cos(k\theta_w)$ .  
940 Strictly speaking,  $\delta\phi(r_w, \theta_w)$  can be separated into the  
941 product of two functions of  $r_w$  and  $\theta_w$  only if all func-  
942 tions  $c_k(r_w)$  are proportional to each other. However,  
943 if these functions decay similarly with the distance to  
944 the wall  $r_w$ , the proportionality assumption and hence  
945 the product assumption would be only weakly invalid,  
946 as illustrated in the practical examples presented above.  
947 In fact, we generally expect only a few Fourier modes  
948 to be relevant in the above expansion (in [12], it was

found that only 1 or 2 non-zero modes were enough to describe all interactions; see next section). Hence, the product assumption can be only severely invalidated if the few relevant functions  $c_k(r_w)$  have very different behaviour. However, on general biological grounds, we expect that all these functions characterizing the interaction with the wall should decay smoothly with the distance  $r_w$ . Similarly, for the corresponding functions describing the attractive interaction between two fish, we expect them to first increase with distance between the fish before decaying at larger range. Hence, even if these few functions are not strictly proportional to each other, their similar anticipated behaviour should lead to effective product interactions grasping the most important features of the actual interactions.

In conclusion of this section, the product assumption allows for an efficient method to reconstruct the individual one-variable interaction functions, only requiring a moderate amount of experimental data. Even if the product assumption is not strictly verified by the actual interactions, general biological considerations ensure that the product interactions would still describe their main features. In any case, the product form offers an explicit representation of the interactions, separating the different contributions (interaction with the wall, attraction/repulsion and alignment between individuals). The reconstructed one-variable interaction functions can be fitted by simple analytic forms (see next section) and the full interaction functions can then be straightforwardly implemented in an explicit and concise model whose predictions can be checked against experimental results, hence providing a further validation of the product assumption. In addition to ultimately producing models for the dynamics of animal groups, the simple and explicit form of the interactions allows for a precise analysis of the behavioural interactions at play in the system, and in particular, for the disentangling of their different components.

## 4 Extraction and comparison of social interactions in different species of fish

The model proposed in Sec. 3 is based on the assumption that social interactions are combined in an additive form (see step 3, Fig. 6). In Eq. (2), two functions  $\delta\phi_{\text{Att}}$  and  $\delta\phi_{\text{Ali}}$  were introduced to account for the attraction and alignment interactions, for which no *a priori* assumptions were made except that their dependence on the state variables  $d_{ij}$ ,  $\psi_{ij}$  and  $\phi_{ij}$  is decoupled; see Eqs. (5)-(6). Steps 4.1 to 4.3 in Fig. 6 provide us with functional forms, but do not determine which functional form corresponds to which kind of interaction.

To do that, the analysis of the relevant observables (step 2 in Fig. 6) is used to say that  $\delta\phi$  must change sign when  $\psi_{ij}$  and  $\phi_{ij}$  both change sign, and that, consequently, the same must happen for  $\delta\phi_{\text{Att}}$  and  $\delta\phi_{\text{Ali}}$ .

This way, the parity of the angular components of the interaction functions is univocally determined. Thus, to have an interaction of attraction, the fish must turn left (resp. right) if its neighbour is on its left (resp. right); that is,  $\delta\phi > 0$  if  $\psi_{ij} > 0$  (resp.  $\delta\phi < 0$  if  $\psi_{ij} < 0$ ). Assuming perfect left/right symmetry, this exactly means that  $\delta\phi_{\text{Att}}$  must be an odd function of  $\psi_{ij}$ , and thus an even function of  $\phi_{ij}$ , provided fish do not have side preferences to turn (that is, fish do not *prefer* turning left to turning right). Similarly, to have an interaction of alignment, the fish must turn left when the relative heading of its neighbour is turned to the left, and turn right if it is turned to the right; that is,  $\delta\phi_{\text{Ali}}$  must be an odd function of  $\phi_{ij}$ , and thus an even function of  $\psi_{ij}$ .

This allows us to rewrite the interaction functions from Eqs. (5)-(6) as follows:

$$\delta\phi_{\text{Att}}(d_{ij}, \psi_{ij}, \phi_{ij}) = F_{\text{Att}}(d_{ij}) O_{\text{Att}}(\psi_{ij}) E_{\text{Att}}(\phi_{ij}), \quad (23)$$

$$\delta\phi_{\text{Ali}}(d_{ij}, \psi_{ij}, \phi_{ij}) = F_{\text{Ali}}(d_{ij}) E_{\text{Ali}}(\psi_{ij}) O_{\text{Ali}}(\phi_{ij}), \quad (24)$$

where function names “O” and “E” stand for “odd” and “even” respectively. The six unknown interaction functions are then tabulated on a grid with typically 30 boxes leading to effectively 180 fitting parameters, a number much smaller than the number of kicks experimentally recorded ( $2 \times 10^5$  for *H. rhodostomus* and  $4 \times 10^4$  in *D. rerio*). Ultimately, these fitting parameters are determined by minimizing the corresponding error function (see Sec. 3).

Fig. 14 shows the social interaction functions reconstructed with the procedure described in the previous section in the case of two *H. rhodostomus* (Panels ABC) and two *D. rerio* (Panels DEF) swimming in circular arenas. In both species, attraction and alignment have been detected and, although some important differences can be observed, each kind of interaction has essentially a similar shape and intensity, although the interaction ranges are very different. Hence, we used the analytical expressions introduced in [12] for *H. rhodostomus* (solid lines in Fig. 14) to fit the discrete values of *D. rerio* extracted from the experimental data in step 4 of our method (Fig. 6), that is, with the procedure described in Section 3 (points in Fig. 14).

These expressions are, for the intensity of the social interaction of attraction and alignment, as follows:

$$F_{\text{Att}}(d) = \gamma_{\text{Att}} \left( \frac{d}{d_{\text{Att}}} - 1 \right) \frac{1}{1 + \left( \frac{d}{l_{\text{Att}}} \right)^2} \quad (25)$$

$$F_{\text{Ali}}(d) = \gamma_{\text{Ali}} \frac{d}{d_{\text{Ali}}} \exp \left[ - \left( \frac{d}{l_{\text{Ali}}} \right)^2 \right], \quad (26)$$

where the values of the parameters depend on the species and on the size of the arena. Note that the expression of the intensity of the alignment has been simplified with respect to the one obtained in [12] and now has one less parameter. Here,  $\gamma_{\text{Att}}$  and  $\gamma_{\text{Ali}}$  are the (dimensionless) intensities of the attraction and alignment interactions,  $d_{\text{Att}}$  is the distance below which attraction changes sign and becomes repulsion,  $l_{\text{Att}}$  and

1051  $l_{\text{Ali}}$  are the ranges of each interaction (the higher the  
1052 value, the longer the range), and  $d_{\text{Ali}} = 1 \text{ BL}$  is a  
1053 characteristic length used to make  $\gamma_{\text{Ali}}$  dimensionless.  
1054 Having dimensionless factors  $\gamma_{\text{Att}}$  and  $\gamma_{\text{Ali}}$  allows direct  
1055 comparison between intensities of social interactions  
1056 and with the intensity of the interaction with obstacles  
1057  $\delta\phi_{\text{w}}$  and the spontaneous decision term  $\delta\phi_{\text{R}}$  in  
1058 Eq. (1). Table 1 shows the parameter values correspond-  
1059 ing to each species. Note that the values for *H. rhodostomus*  
1060 have been adapted from those in [12], according  
1061 to the slightly different expression used here.

	<i>H. rhodostomus</i>	<i>D. rerio</i>
$R$ (m)	0.25	0.29
$R/\text{BL}$	8.33	6.44
$\gamma_{\text{Att}}$	0.124	0.42
$d_{\text{Att}}$ (m)	0.03	0.015
$l_{\text{Att}}$ (m)	0.193	0.042
$\gamma_{\text{Ali}}$	0.092	0.32
$d_{\text{Ali}}$ (m)	0.03	0.045
$l_{\text{Ali}}$ (m)	0.16	0.1

1063 Table 1. Parameter values for *H. rhodostomus* and  
1064 *D. rerio* in circular arenas of radius 0.25 m and 0.29 m  
1065 respectively.

Regarding the normalized angular functions for  
*D. rerio*, we found the following expansions (with at  
most 2 Fourier modes in addition to the trivial zero  
mode):

$$\begin{aligned}
 O_{\text{Att}}(\psi) &= 1.66 \sin(\psi)[1 - 0.77 \cos(\psi) + 0.6 \cos(2\psi)], \\
 E_{\text{Att}}(\phi) &= 0.81[1 - 0.95 \cos(\phi) + 0.12 \cos(3\phi)], \\
 E_{\text{Ali}}(\psi) &= 0.54[1 + \cos(\psi) - 2 \cos(2\psi)], \\
 O_{\text{Ali}}(\phi) &= 1.53 \sin(\phi)[1 + 0.24 \cos(\phi)].
 \end{aligned}$$

1066 Following the step 5 of our methodology (Fig. 6),  
1067 these analytical expressions should be implemented in  
1068 the model introduced in Section 3. Then, numerical  
1069 simulations of the model should be performed and  
1070 compared with the known experiments, and predictions  
1071 should be made, that must be verified *a posteriori*. This  
1072 corresponds to the step 6 of our methodology (Fig. 6);  
1073 it was done for *H. rhodostomus* in [12], and will be done  
1074 elsewhere for *D. rerio* and other species.

1075 Comparing the interaction functions found for both  
1076 species, we observe that they have a similar shape, espe-  
1077 cially the angular functions (see the angular functions of  
1078 attraction in Fig. 14B, for which one can use the same  
1079 function of the angle of perception  $\psi_{ij}$ ). The intensi-  
1080 ties of attraction and alignment have the same order of  
1081 magnitude in both species: the maximum of the attrac-  
1082 tion is around 0.4 in both species, and the maximum of  
1083 the alignment is around 0.2–0.3 (Fig. 14A).

1084 The most important difference between the two  
1085 species is that the range of the interactions is much  
1086 larger in *H. rhodostomus* than in *D. rerio*: in *H. rhodos-*  
1087 *tomus*, the maximum intensities of attraction and align-  
1088 ment are around 7 BL and 3.5 BL respectively, while in

1089 *D. rerio* these maxima are both near 1.5 BL. Moreover,  
1090 the intensity of these interactions decays more rapidly  
1091 in *D. rerio* than in *H. rhodostomus*, especially with re-  
1092 spect to the fish body length (in *D. rerio*, the alignment  
1093 intensity is zero beyond 5 BL  $\approx 22.5$  cm, but it is still  
1094 noticeable at a distance of 10 BL  $\approx 30$  cm in *H. rhodos-*  
1095 *tomus*). Attraction almost always dominates alignment  
1096 in *D. rerio* (the intersection of the red and blue lines  
1097 in Fig. 14A is at around 0.5 BL), while, in *H. rhodos-*  
1098 *tomus*, alignment was found to dominate attraction at  
1099 short distances (under 2.5 BL).

1100 In *H. rhodostomus*, a fish  $i$  is subject to a stronger  
1101 attraction when the other fish  $j$  is at its right or left  
1102 side ( $O_{\text{Att}}$  reaches its highest values when  $\psi_{ij} \approx \pm 90^\circ$ ;  
1103 see Fig. 14B) and moves more or less perpendicular to  
1104 it ( $E_{\text{Att}}$  is higher when  $\phi_{ij} \approx 90\text{--}100^\circ$ ; see Fig. 14C).  
1105 Alignment is stronger when the other fish is in front  
1106 ( $E_{\text{Ali}}$  is higher when  $|\psi_{ij}| < 80^\circ$ ). In *D. rerio*, attrac-  
1107 tion is stronger when the other fish is clearly behind  
1108 the focal fish ( $O_{\text{Att}}$  is higher when  $\psi_{ij} \approx \pm 135^\circ$ ) and  
1109 moves in the opposite direction ( $E_{\text{Att}}$  is higher when  
1110  $|\phi_{ij}| > 100^\circ$ ; see Figs. 14EF).

1111 In both species, the strength of the alignment van-  
1112 ishes when fish are already almost aligned ( $O_{\text{Ali}} \approx 0$   
1113 when  $|\phi_{ij}| < 30^\circ$ ; see panels C and F in Fig. 14). In  
1114 *D. rerio*, the strength of the alignment is active essen-  
1115 tially when both fish are perpendicular to each other  
1116 (the high intensity of  $|O_{\text{Ali}}| \approx 1.5$  is reached when  
1117  $\phi_{ij} \approx \pm 85^\circ$ ) and the focal fish has its neighbour at  
1118 one of its sides ( $E_{\text{Ali}}$  is peaked at  $\psi_{ij} \approx \pm 85^\circ$ ), while  
1119 in *H. rhodostomus*, the ranges of interaction both in  
1120 the angle of relative heading and in the angle of per-  
1121 ception of the neighbour are much wider: alignment  
1122 is active when the neighbour is ahead of the focal  
1123 fish ( $|\psi_{ij}| < 100^\circ$ ) and fish are simply slightly aligned  
1124 ( $45^\circ < |\phi_{ij}| < 135^\circ$ ).

1125 In summary, when swimming in pairs, *H. rhodosto-*  
1126 *mus* interact in a much wider range of situations than  
1127 *D. rerio*. This is true with respect to the three state  
1128 variables of a focal fish: 1) the distance  $d_{ij}$  at which  
1129 both attraction and alignment interactions are active is  
1130 much larger in *H. rhodostomus* than in *D. rerio*; 2) the  
1131 zone around a focal fish where the strength of the inter-  
1132 actions with a neighbour is important is much wider in  
1133 *H. rhodostomus* than in *D. rerio*, and 3) the same is true  
1134 for the range of relative headings for which the intensity  
1135 of interactions between fish is not negligible. Finally,  
1136 the maximum intensity of the interaction is similar in  
1137 both species, although high intensity values are reached  
1138 in a much wider range of situations in *H. rhodostomus*  
1139 than in *D. rerio*.

## 1140 5 Discussion and conclusions

1141 Behavioural biology has recently become a “big-data  
1142 science” mainly supported by the advances in imaging  
1143 and tracking techniques. These new tools have revolu-  
1144 tionized the observation and quantification of individual

and collective animal behaviour, improving to unprecedented levels the variety and precision of available data [9, 26, 27, 28]. As the access to large volumes of data is gradually stepping animal behaviour research into a new era, there is also a growing need for understanding interactions between individuals and the collective properties that emerge from these interactions. Animal societies are complex systems whose properties are not only qualitatively different from those of their individual members, but whose behaviours are impossible to predict from a prior knowledge of individuals [3, 4]. However, understanding how the interactions between individuals in swarms of insects, schools of fish, flocks of birds, herds of ungulates, or human crowds give rise to the “collective level” properties requires the development of mathematical models. These models allow to investigate how complex processes are connected, to systematically analyse the impact of perturbations on collective behaviour (*e.g.*, when a predator is detected in the neighbourhood), to develop hypotheses to guide the design of new experimental tests, and ultimately, to assess how each biological variable contributes to the emergent group level properties.

We have presented a general methodology which leads to the measurement of the social interactions (defined in the framework of a model) from a set of individual trajectories. This procedure, illustrated here on two different species of fish, can be similarly applied on any set of trajectories of other organisms, including humans [29, 30]. Once the experimental trajectories have been obtained, the extraction of interaction functions only makes sense after they have been defined in the framework of a general model for the equation of motion of an individual interacting with its environment (*i.e.*, the obstacles and another individual). The model should involve the relevant variables regarding the interaction with obstacles (distance to the nearest wall, angle of the velocity with respect to the normal to the wall...) and another individual (distance between individuals, relative velocities, viewing angle...). In addition, each interaction components (repulsion, attraction, alignment...) is reasonably assumed to contribute additively: for instance, the influence of the wall and another individual on the focal individual is the independent sum of the two corresponding interactions. More importantly, the central hypothesis of our approach consists in assuming that the contribution of each interaction can be adequately described by a product of unknown single-variable functions of the relevant variables, or of a combination of these variables. The structure of the model is obviously also constrained by the considered species, their motion mode, and their anticipated interactions. For instance, for the fish species that have a burst-and-coast swimming mode, the dynamical model is intrinsically discrete in time and returns the angle change of an individual after each kick. For humans or some other fish species with a smooth swimming mode, a continuous-time model is necessary.

The unknown interaction functions defined in the

structure of the model are not constrained, and the aim of the extraction procedure (step 4 in Fig. 6) is to measure them, without any *a priori* assumption about their form or intensity. In order to achieve that, each unknown single-variable interaction function is tabulated on a one-dimensional grid, and their values at each grid point are the fitting parameters. These parameters are then determined by minimizing the mean quadratic error between the prediction of the model and the experimental angle changes after a kick (in the case of discrete dynamics) or the experimental acceleration (in the case of a continuous time dynamics). This minimization process (Box 1) is achieved by solving, for instance with an iterative method (Box 2), the equations expressing the vanishing of the partial derivatives of the error with respect to each fitting parameters, as we have done here, or by gradient descent methods. Once the interaction functions have been obtained on their respective grid, they are fitted and represented by simple analytical forms that best capture their general shape (Fig. 14).

The original general equation of motion is now complemented by these explicit interaction functions, leading to a concise and explicit model, which can be straightforwardly implemented numerically and can even be studied mathematically. The ability of the model to reproduce the experimental results and even to predict the behaviour of individuals in other situations not yet investigated experimentally can then be assessed. The original model can also be extended and the full extraction procedure repeated if an important feature appears to be missing. In particular, quantities like the probability distributions of the distance to the wall, of the distance between two individuals, of the angle between the two velocities or between the velocity and the normal to the wall, as well as other observables, permit to assess the predictive power of the model. Note that the interaction functions appearing in the model are determined by finding the best equation of motion describing the instantaneous decisions of the individuals. It is by no means trivial that this is enough for the resulting dynamical model to be able to reproduce observable quantities measured after averaging over many trajectories, or to predict the behaviour of individuals in different experimental conditions. A model able to achieve this certainly provides a convincing indication that its original and general structure and its extracted interaction functions properly represent and describe the behaviour and motion of the studied species.

The main limitations of our methodology lie in the reasonable assumption of additive contributions for the different interactions, and more critically, in assuming that each of these interactions is the product of single-variable interaction functions. This is the cost to pay for only involving a limited number of fitting parameters, yet capturing a large part of the complex structure of these interactions, and also for ultimately obtaining a concise, explicit, and exploitable model.

To address these issues, we are planning in the near

future to test our models on a robotic platform [31]. Such a platform could also allow us to study bidirectional interactions between robots reproducing the trajectories generated by our models and real fish, to obtain a more representative validation of our models of interactions [32].

Yet, the advantages and benefits of our approach are numerous. First, the number of fitting parameters (typically 30 for each of the typically 5-8 interaction functions), although apparently large (a total of typically 150-300 parameters), is in general much smaller than the number of data points in the available experimental trajectories (typically  $10^4$  for human groups [29], and  $10^5$  for fish [12]). In comparison, a complete force map in typically 5 or more dimensions (one dimension per relevant variable) on a mesh involving  $30^5$  boxes, with enough data points in each of them, would require millions if not billions of experimental data points, which is in general impossible to achieve. Two-dimensional force maps obtained after projection (*i.e.*, averaging on the other  $5 - 2 = 3$  variables), although less noisy than the original force map, cannot be exploited to build a model, since it can be shown that they are strongly affected by the existing correlations between these variables along actual trajectories. For instance, if a fish is very close to a wall, there is a high probability that it swims parallel to the wall, so that its distance from the wall is strongly correlated with its heading angle. On the other hand, it can be shown [12] that our methodology to extract interaction functions is not at all affected by the likely correlations present in the system, and actually exploit them.

Our method is also robust with respect to the presence of noise in the data (intrinsic behavioural noise or unwanted experimental noise), and can actually be used to measure the spontaneous fluctuations of the speed and heading angle of the individuals [12, 30]. Moreover, the extraction of interaction functions requires very limited computing power, being obtained within a few seconds on a standard workstation. Ultimately, the analysis of the resulting interaction function allows to make general qualitative conclusions about the interaction at play for the considered species. Force maps can also help in this analysis, but our approach allows for an even finer analysis thanks to the disentangling of interactions by means of separate and explicit interaction functions, instead of mere projected colour maps affected in an uncontrolled manner by the inherent correlations present in the system and the mixing of the different interaction contributions.

More importantly, our methodology ultimately leads to a concise and explicit model which can be exploited to understand and explain diverse experimental features and various forms of collective behaviour, and which has a predictive power, while force maps cannot be directly exploited to build such explicit models. Note that the structure of the model should be robust for different species having comparable motion mode. For instance, this structure is the same for *H. rhodostomus* and *D. rerio*

(or for any species with a burst-and-coast or run-and-tumble motion mode), and the behavioural differences between the species is solely and fully encoded in their different measured interaction functions.

In the specific case of *H. rhodostomus* and *D. rerio*, we have also found that the interaction functions characterizing their interaction with the wall are very similar (see [12] for *H. rhodostomus*), which explains their common tendency to swim close to the wall, especially when a fish swims alone in a tank. However, even if the interaction functions describing the repulsion/attraction and alignment interactions between two fish have a similar general structure and shape for both species (Fig. 14), the range of the attraction and alignment interactions is much shorter for *D. rerio*. In addition, the intensity of both interactions in *D. rerio* is strongly reduced when the focal fish is behind and hence follows the other fish (*i.e.*, when  $|\psi_{ij}| < 30^\circ$ ). Both features contribute to a much weaker coordination of the motion in groups of two fish in *D. rerio*, compared to *H. rhodostomus*, which can already be qualitatively noticed by observing recorded trajectories. This weaker coordination is quantitatively illustrated by the probability distribution of the distance between two fish (panels A and D in Fig. 14), which is wider for *D. rerio* and extends up to much larger distances. Moreover, the probability distribution of the heading angle difference between the two fish is less peaked near  $\phi_{ij} = 0$  in *D. rerio* (Fig. 14F). Finally, the weaker coordination observed in *D. rerio* results in a similar behaviour for the geometrical leader and follower, whereas the leader and follower have a very distinct behaviour in *H. rhodostomus*.

## Materials and methods

**Ethics.** Experiments with *H. rhodostomus* have been approved by the Ethics Committee for Animal Experimentation of the Toulouse Research Federation in Biology No. 1 and comply with the European legislation for animal welfare. Experiments with *D. rerio* were conducted under authorization approved by the state ethical board of the Department of Consumer and Veterinary Affairs of the Canton de Vaud (SCAV) of Switzerland (authorization No. 2778). During the experiments, no mortality occurred.

**Study species.** *H. rhodostomus* were purchased from Amazonie Labège (<http://www.amazonie.com>) in Toulouse, France. Fish were kept in 150L aquariums on a 12:12 hour, dark:light photoperiod, at  $26.8^\circ\text{C}$  ( $\pm 1.6^\circ\text{C}$ ) and were fed *ad libitum* with fish flakes. The average body length of the fish used in the experiments was 31 mm. Wild-type *D. rerio* with short fins (AB strain) were acquired in a number of 60 from a pet shop, and stored in a 60-litre aquarium. The average body length of the fish used in the experiments was approximately 4.5 cm in length. The water in the housing aquarium was kept at a temperature of  $26^\circ\text{C}$ . The fish

were fed once per day with commercial food between 16:00 and 18:00. Additionally, we opted to use enrichment for the aquarium in the form of plastic plants, Cladophora, gravel, rocks, and aquatic snails.

### Experimental procedures and data collection.

The experimental tank ( $120 \times 120 \text{ cm}^2$ ) used to investigate swimming behaviour in *H. rhodostomus* was made of glass and was set on top of a box to isolate fish from vibrations. The setup, placed in a chamber made by four opaque white curtains, was surrounded by four LED light panels giving an isotropic lighting. A circular tank of radius  $R = 25 \text{ cm}$  was set inside the experimental tank filled with 7 cm of water of controlled quality (50% of water purified by reverse osmosis and 50% of water treated by activated carbon) heated at  $26.1^\circ\text{C}$  ( $\pm 0.3^\circ\text{C}$ ). Reflections of light due to the bottom of the experimental tank are avoided thanks to a white PVC layer. Each trial started by setting two fish randomly sampled from their breeding tank into a circular tank. Fish were let for 10 minutes to habituate before the start of the trial. A trial consisted in one or three hours of fish freely swimming (*i.e.*, without any external perturbation) in a circular tank. A total of 16 trials were performed. Fish trajectories were recorded by a Sony HandyCam HD camera filming from above the set-up at 50 Hz (50 frames per second) in HDTV resolution ( $1920 \times 1080\text{p}$ ).

In the case of *D. rerio*, the experimental setup had dimensions of  $100 \times 100 \times 25 \text{ cm}^3$ , inside which a circular tank of radius  $R = 29 \text{ cm}$  was placed. The sides and bottom part of the tank were covered with a Teflon plate to avoid reflections. Furthermore, the setup was confined behind white sheets to isolate the fish from external stimuli in the room, while also maintaining a consistent lighting environment inside the setup bounds. A uniform luminosity for the room was provided by four 110 watt fluorescent lamps placed at each of the four sides of the tank. Prior to placing fish in the experimental setup, we ensured that the height of the water was 6 cm. Then, 10 videos of 70 min duration each were recorded in the circular tank. Subsequently, a group of fish was randomly selected and caught from the rearing tanks to participate in the experiment. A pair of fish was then chosen and placed in the setup. The fish were allowed to habituate for 5 min before starting the

70 min long recording. After a single experiment was completed, the fish were returned to the original rearing tank without being re-inserted in the selected group (*i.e.*, no individual was used twice in the same day). The positions of fish on each frame were tracked with idTracker 2.1 [8]. Time series of positions were converted from pixels to meters and the origin of the coordinate system was set to the centre of the ring-shaped tank. Tracking errors (approx. 20% of the data) were corrected and instances where at least one fish moves less than 0.5 body length per second during 4 seconds were removed. More than 8 hours remained during which one fish can kick.

We provide the dataset of each species in Figshare: <https://doi.org/10.6084/m9.figshare.9933356.v1>.

**Segmentation of trajectories.** Instants of kicks were identified as local minima of the velocity preceding local maxima (which are more easy to identify), along time intervals  $[t - t_w, t + t_w]$ , where  $t_w$  is a time window of 0.32 seconds. Trajectories are thus considered as a sequence of kicks. Fish almost never kick at the same time (asynchronous kicks); the position of the other fish is calculated at each instant of kick of the focal fish by simple linear interpolation.

**FORTRAN code.** We provide the code to perform the analysis described in Sec. 3.2, together with a simple script for quick visualisation of the result in the `gnuplot` environment and the required data, in Figshare: <https://doi.org/10.6084/m9.figshare.11777325>.

## Acknowledgements

This work was supported by the Germaine de Staël project No. 2019-17 and the Swiss National Science Foundation project “Self-Adaptive Mixed Societies of Animals and Robots”, Grant No. 175731. R.E. was supported by Marie Curie Core/Program Grant Funding, Grant No. 655235–SmartMass. V.L. was supported by doctoral fellowships from the scientific council of the Université Paul Sabatier. G.T. gratefully acknowledges the Indian Institute of Science to serve as Infosys visiting professor at the Centre for Ecological Sciences in Bengaluru. We thank Gérard Latil for technical assistance.

## References

- [1] Kuniyiko Kaneko. Life: An Introduction to Complex Systems Biology. Understanding Complex Systems. Springer-Verlag, Berlin, 2006.
- [2] Ricard V Solé. Phase Transitions. Princeton University Press, Princeton, NJ, 2011.
- [3] Scott Camazine, Jean-Louis Deneubourg, Nigel Franks, James Sneyd, Guy Theraulaz, and Eric Bonabeau. Self-Organization in Biological Systems. Princeton studies in complexity. Princeton University Press, Princeton, NJ, 2001.
- [4] David JT Sumpter. Collective Animal Behavior. Princeton University Press, Princeton, NJ, 2010.
- [5] Christian Jost, Julie Verret, Eric Casellas, Jacques Gautrais, Mélanie Challet, Jacques Lluc, Stéphane
- [6] E. Casellas, J. Gautrais, R. Fournier, S. Blanco, M. Combe, V. Fourcassié, G. Theraulaz, and C. Jost. From in-

- dividual to collective displacements in heterogeneous environments. Journal of Theoretical Biology, 250(3):424–434, February 2008.
- [7] Kristin Branson, Alice A Robie, John Bender, Pietro Perona, and Michael H Dickinson. High-throughput ethomics in large groups of *Drosophila*. Nature Methods, 6:451, May 2009.
- [8] Alfonso Pérez-Escudero, Julián Vicente-Page, Robert C Hinz, Sara Arganda, and Gonzalo G. de Polavieja. idTracker: Tracking individuals in a Group by Automatic Identification of Unmarked Animals. Nature Methods, pages 743–748, 2014.
- [9] Anthony I. Dell, John A. Bender, Kristin Branson, Iain D. Couzin, Gonzalo G. de Polavieja, Lucas P. J. J. Noldus, Alfonso Pérez-Escudero, Pietro Perona, Andrew D. Straw, Martin Wikelski, and Ulrich Brose. Automated Image-Based Tracking and Its Application in Ecology. Trends in Ecology Evolution, 29(7):417 – 428, 2014.
- [10] D.J. Anderson and Perona P. Toward a science of computational ethology. Neuron, 1(84).
- [11] F. Romero-Ferrero, M.G. Bergomi, R.C. Hinz, F.J.H. Heras, and G.G. de Polavieja. idtracker.ai: tracking all individuals in small or large collectives of unmarked animals. Nature Methods, (16).
- [12] Daniel S. Calovi, Alexandra Litchinko, Valentin Lecheval, Ugo Lopez, Alfonso Pérez Escudero, Hugues Chaté, Clément Sire, and Guy Theraulaz. Disentangling and Modeling Interactions in Fish with Burst-and-Coast Swimming Reveal Distinct Alignment and Attraction Behaviors. PLoS Computational Biology, 14(1):1–28, January 2018.
- [13] Adam K. Zienkiewicz, Fabrizio Ladu, David A. W. Barton, Maurizio Porfiri, and Mario Di Bernardo. Data-Driven Modelling of Social Forces and Collective Behaviour in Zebrafish. Journal of Theoretical Biology, 443:39 – 51, 2018.
- [14] U. Lopez, J. Gautrais, I. D. Couzin, and G. Theraulaz. From Behavioural Analyses to Models of Collective Motion in Fish Schools. Interface Focus, 2(6):693–707, December 2012.
- [15] J. E. Herbert-Read. Understanding How Animal Groups Achieve Coordinated Movement. Journal of Experimental Biology, 219(19):2971–2983, 2016.
- [16] Liu Lei, Ramón Escobedo, Clément Sire, and Guy Theraulaz. Computational and robotic modeling reveal parsimonious combinations of interactions between individuals in schooling fish. PLoS Computational Biology (in press), 2020.
- [17] V. Mwaffo, R. P. Anderson, S. Butail, and M. Porfiri. A Jump Persistent Turning Walker to Model Zebrafish Locomotion. Journal of The Royal Society Interface, 12(102):20140884–20140884, January 2015.
- [18] Yael Katz, Kolbjørn Tunstrøm, Christos C. Ioannou, Cristián Huepe, and Iain D. Couzin. Inferring the Structure and Dynamics of Interactions in Schooling Fish. Proceedings of the National Academy of Sciences, 108(46):18720–18725, 2011.
- [19] James E. Herbert-Read, Andrea Perna, Richard P. Mann, Timothy M. Schaerf, David JT Sumpter, and Ashley JW Ward. Inferring the Rules of Interaction of Shoaling Fish. Proceedings of the National Academy of Sciences, 108(46):18726–18731, 2011.
- [20] S. Lee, G. Wolberg, and S. Y. Shin. Scattered data interpolation with multilevel B-splines. IEEE Transactions on Visualization and Computer Graphics, 3(3):228–244, July 1997.
- [21] Sebastian Weitz, Stéphane Blanco, Richard Fournier, Jacques Gautrais, Christian Jost, and Guy Theraulaz. Modeling Collective Animal Behavior with a Cognitive Perspective: A Methodological Framework. PLoS ONE, 7(6):e38588, June 2012.
- [22] Colin J. Torney, Myles Lamont, Leon Debell, Ryan J. Angohiatok, Lisa-Marie Leclerc, and Andrew M. Berdahl. Inferring the rules of social interaction in migrating caribou. Philosophical Transactions of the Royal Society B: Biological Sciences, 373(1746):20170385, May 2018.
- [23] Vicenç Quera, Elisabet Gimeno, Francesc S. Beltran, and Ruth Dolado. Local interaction rules and collective motion in black neon tetra (*Hyphessobrycon herbertaxelrodi*) and zebrafish (*Danio rerio*). Journal of Comparative Psychology, 133(2):143–155, 2019.
- [24] Ryan Lukeman, Yue-Xian Li, and Leah Edelstein-Keshet. Inferring individual rules from collective behavior. Proceedings of the National Academy of Sciences, 107(28):12576, July 2010.
- [25] Benjamin Pettit, Andrea Perna, Dora Biro, and David J. T. Sumpter. Interaction rules underlying group decisions in homing pigeons. Journal of The Royal Society Interface, 10(89):20130529, December 2013.
- [26] S.E. Roian Egnor and Kristin Branson. Computational analysis of behavior. Annual Review of Neuroscience, 39(1):217–236, 2016. PMID: 27090952.
- [27] Alice A. Robie, Kelly M. Seagraves, S. E. Roian Egnor, and Kristin Branson. Machine vision methods for analyzing social interactions. Journal of Experimental Biology, 220(1):25–34, 2017.
- [28] Tim Gernat, Vikyath D. Rao, Martin Middendorf, Harry Dankowicz, Nigel Goldenfeld, and Gene E. Robinson. Automated monitoring of behavior reveals bursty interaction patterns and rapid spreading dynamics in honeybee social networks. Proceedings of the National Academy of Sciences, 115(7):1433–1438, 2018.
- [29] Bertrand Jayles, Ramón Escobedo, Roberto Pasqua, Christophe Zanon, Adrien Blanchet, Matthieu Roy, Gilles Trédan, Guy Theraulaz, and Clément Sire. Collective information processing in human phase separation. Submitted, 2019.
- [30] Ramón Escobedo, Bertrand Jayles, Gilles Trédan, Matthieu Roy, Roberto Pasqua, Christophe Zanon, Adrien Blanchet, Guy Theraulaz, and Clément Sire. Measuring and modeling individual level interactions in a swarming crowd. In preparation, 2019.
- [31] Frank Bonnet, Philippe Rétoznaz, José Halloy, Alexey Gribovskiy, and Francesco Mondada. Development of a mobile robot to study the collective behavior of zebrafish. In 2012 4th IEEE RAS & EMBS International Conference on Biomedical Robotics and Biomechatronics (BioRob), pages 437–442. Ieee, 2012.

- [32] Vaios Papaspyros, Frank Bonnet, Bertrand Collignon, and Francesco Mondada. Bidirectional interactions facilitate the integration of a robot into a shoal of zebrafish *danio rerio*. PloS one, 14(8):e0220559, 2019.

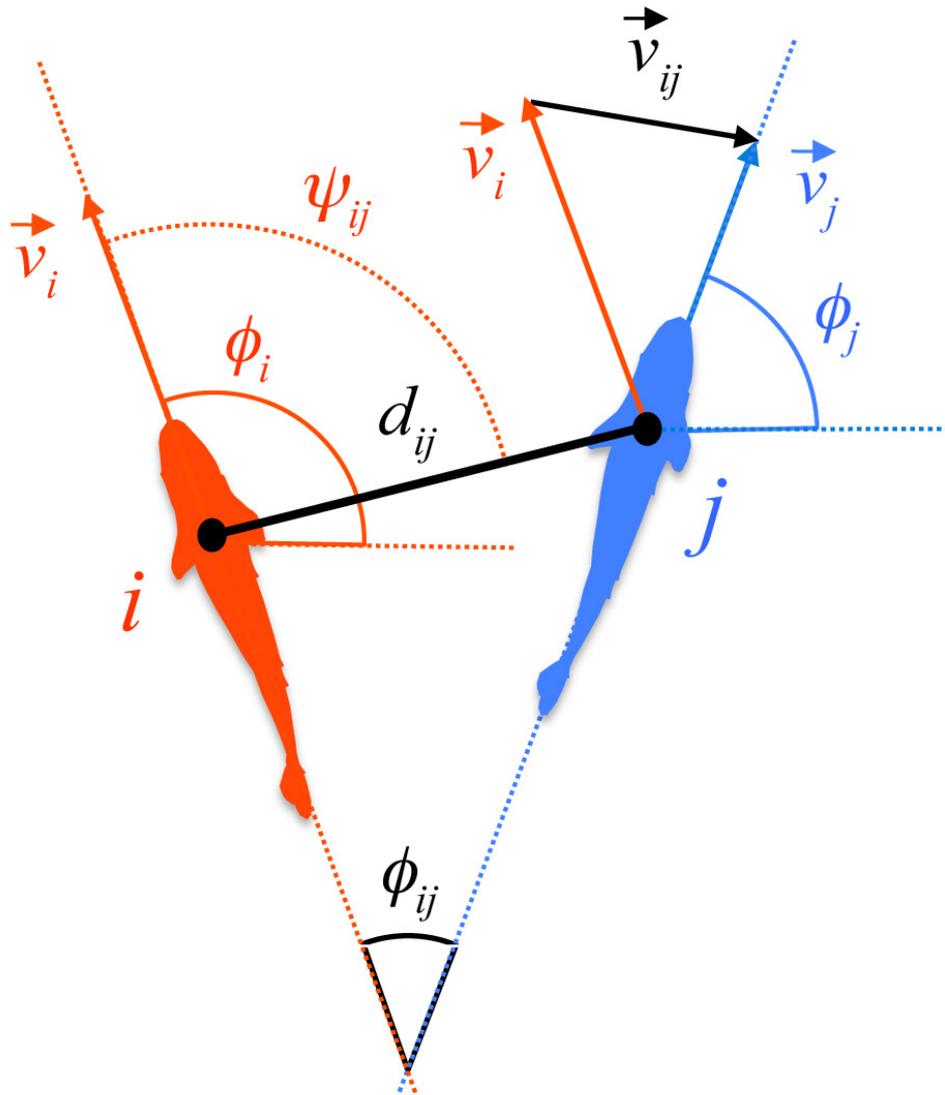


Figure 1: State variables of a focal fish  $i$  with respect to its neighbour  $j$ . Fish position is determined by the position of the centre of mass of the fish (black circles) in a orthonormal system of reference  $Oxy$ .  $d_{ij}$ : distance between fish;  $\vec{v}_{ij} = \vec{v}_j - \vec{v}_i$ : relative velocity of  $j$  with respect to  $i$ ;  $\psi_{ij}$ : viewing angle with which  $i$  perceives  $j$ ;  $\phi_{ij} = \phi_j - \phi_i$ : relative heading of  $j$  with respect to  $i$ . Angles are measured with respect to the horizontal axis of coordinates  $Ox$ ; we use the convention that angles are positive in the counterclockwise direction.

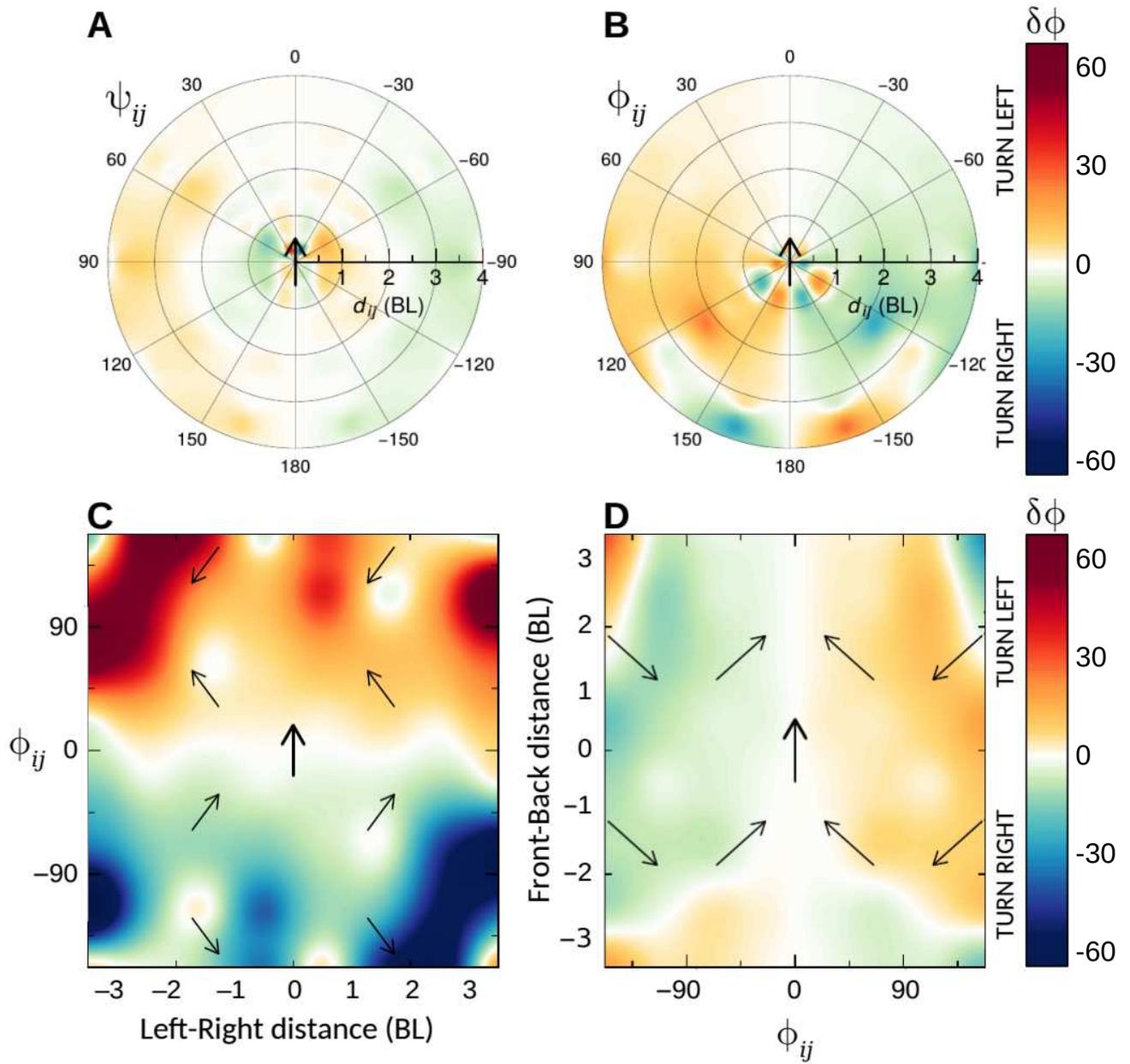


Figure 2: Force maps of *H. rhodostomus* swimming in a circular arena of radius 0.25 m. Heading change  $\delta\phi$  of a focal fish, located at the origin and pointing north in the four panels, as a function of (A) the distance to its neighbour  $d_{ij}$  and the angle of perception of its neighbour  $\psi_{ij}$ , (B) the distance  $d_{ij}$  and its relative heading  $\phi_{ij}$ , (C) the left-right distance  $d_{ij}^{LR}$  and  $\phi_{ij}$ , and (D) the relative heading  $\phi_{ij}$  and the front-back distance  $d_{ij}^{FB}$ . The colour scale shown in the right represents the average value of heading change  $\delta\phi$ . Distances are measured in body lengths (BL), angles in degrees.

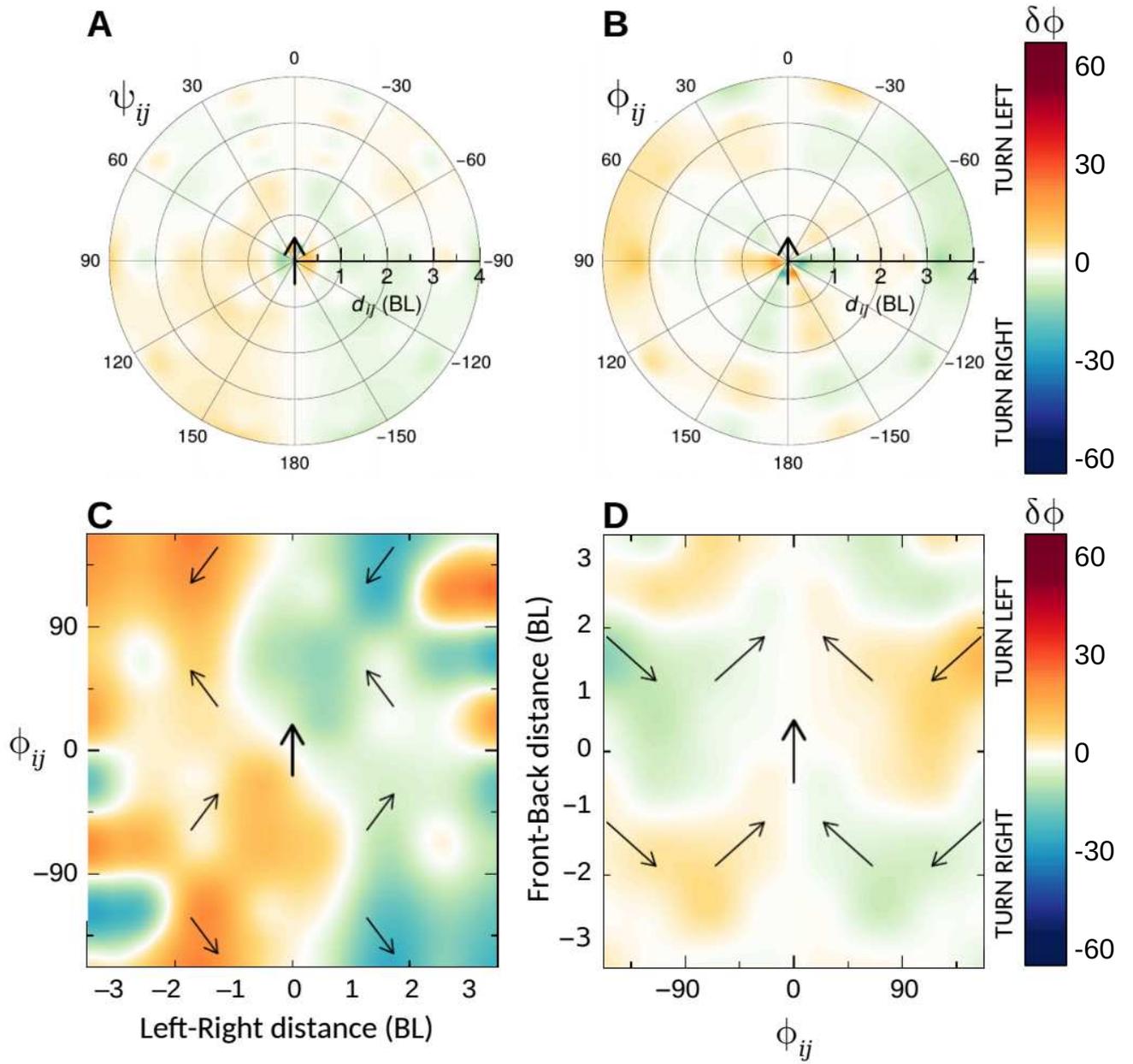


Figure 3: Force maps of *D. rerio* swimming in a circular arena of radius 0.29 m. Heading change  $\delta\phi$  of a focal fish, located at the origin and pointing north in the four panels, as a function of (A) the distance to its neighbour  $d_{ij}$  and the angle of perception of its neighbour  $\psi_{ij}$ , (B) the distance  $d_{ij}$  and its relative heading  $\phi_{ij}$ , (C) the left-right distance  $d_{ij}^{LR}$  and  $\phi_{ij}$ , and (D) the relative heading  $\phi_{ij}$  and the front-back distance  $d_{ij}^{FB}$ . The colour scale shown in the right represents the average value of heading change  $\delta\phi$ . Distances are measured in body lengths (BL), angles in degrees.

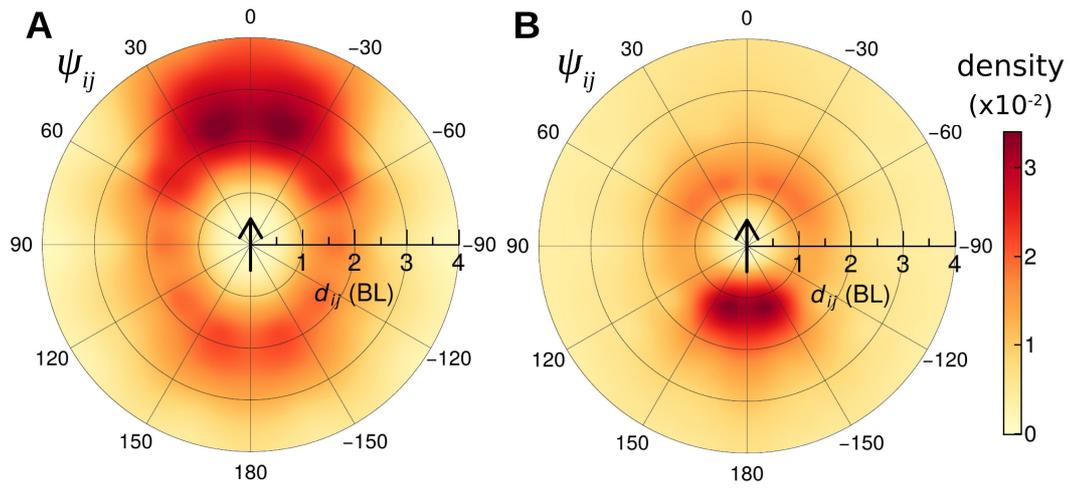


Figure 4: Density maps of the location of a neighbouring fish  $j$  in the system of reference centred on the focal fish  $i$ . (A) *H. rhodostomus* in an arena of radius 0.25 m, and (B) *D. rerio* in the arena of radius 0.29 m, when the focal fish is at least at 2 BL (6 cm) in *H. rhodostomus* and 1 BL (4.5cm) in *D. rerio* from the wall of the tank. The vertical arrow pointing north located at the centre of the coordinate system indicates the position and orientation of the focal fish  $i$ .

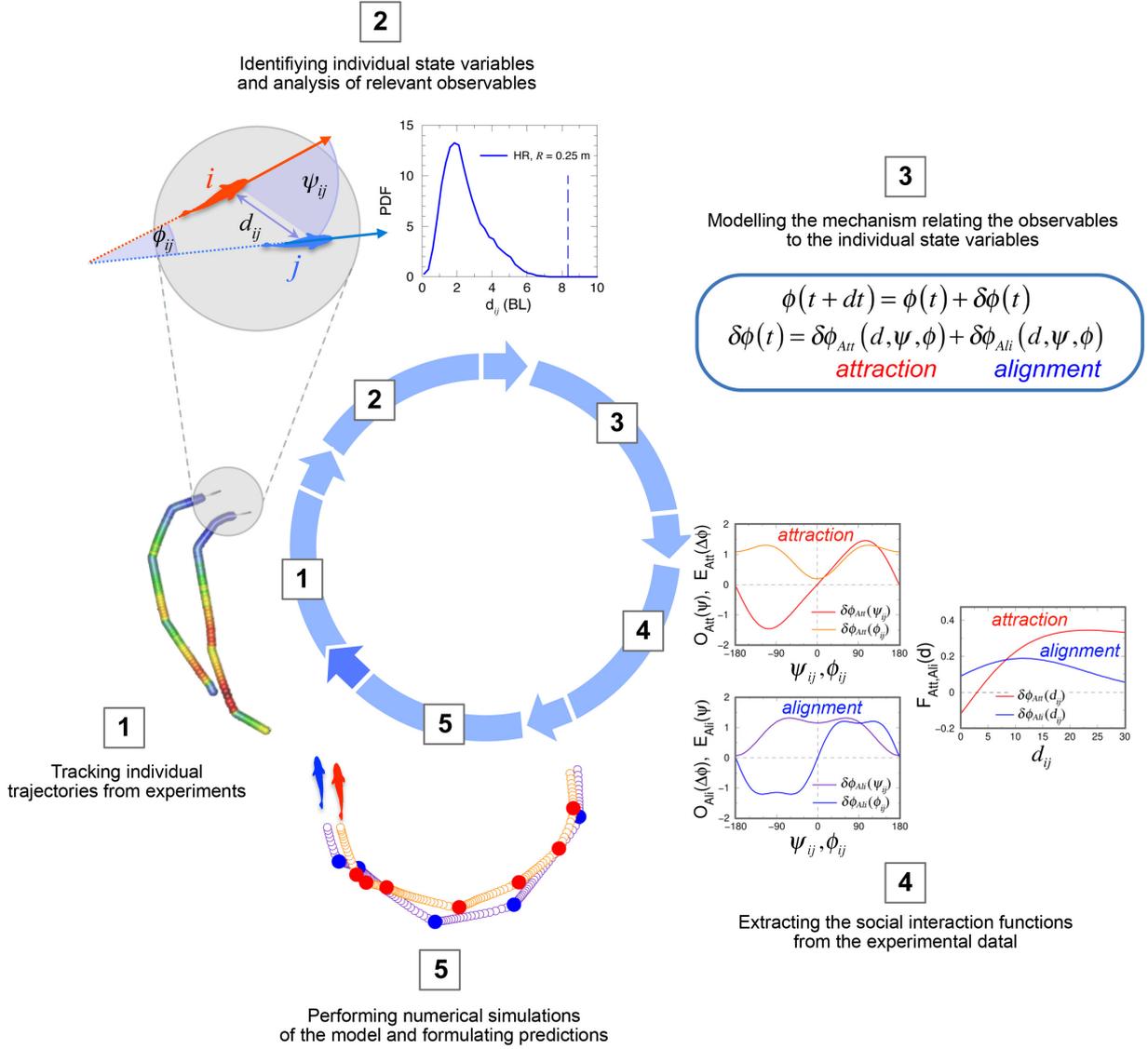


Figure 5: General overview of the methodology. The successive steps carried out in our method are labeled with a number and illustrated with a representative picture: 1: Trajectories of real fish from experimental data; 2: Observables of collective behaviour, the distance  $d_{ij}$  between fish, the relative position  $\phi_{ij}$ , and the relative orientation  $\psi_{ij}$ , and PDF of  $d_{ij}$ ; 3: Model equations of the heading angle changes  $\delta\phi$ ; 4: Extracted interaction functions  $F_{Att}$ ,  $F_{Ali}$ ,  $E_{Att}$ ,  $O_{Att}$ ,  $O_{Ali}$ , and their dependence on  $d_{ij}$ ,  $\psi_{ij}$ , and  $\phi_{ij}$ ; 5: Trajectories resulting from the numerical simulations of the model. The dark blue arrow illustrates the possibility that the cycle starts again, which includes the realisation of a new set of experiments or the design of a new setup, to check the predictions of the model or to deepen our understanding of the mechanisms at work.

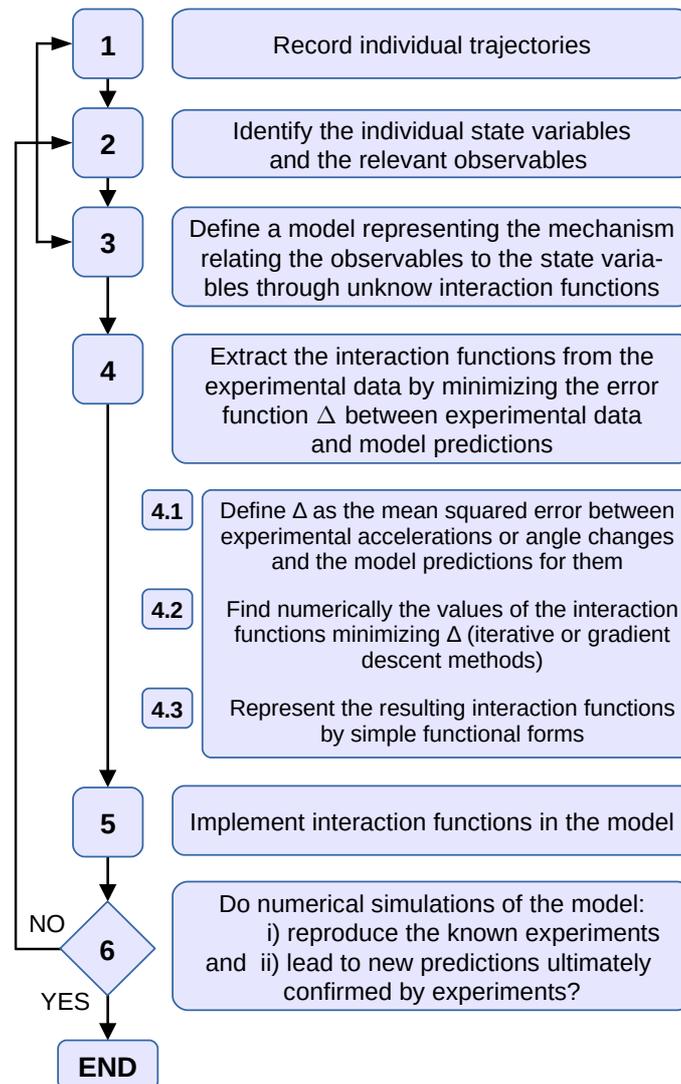


Figure 6: Flowchart of the methodology, with a special emphasis on the procedure of extraction of the interaction functions (step 4). When the test at point 6 is not verified, the cycle starts again at a previous step, depending on the reason for which the test failed: if the model has to be revised (*e.g.*, because the relation between observables and state variables is not well reproduced), then go to step 3; if other state variables or observables have to be considered in the model (*e.g.*, because they may contribute to some observed effect), then go to step 2; if new experiments must be carried out (*e.g.*, because there are not enough data or some important case has not been observed), then go to step 1.

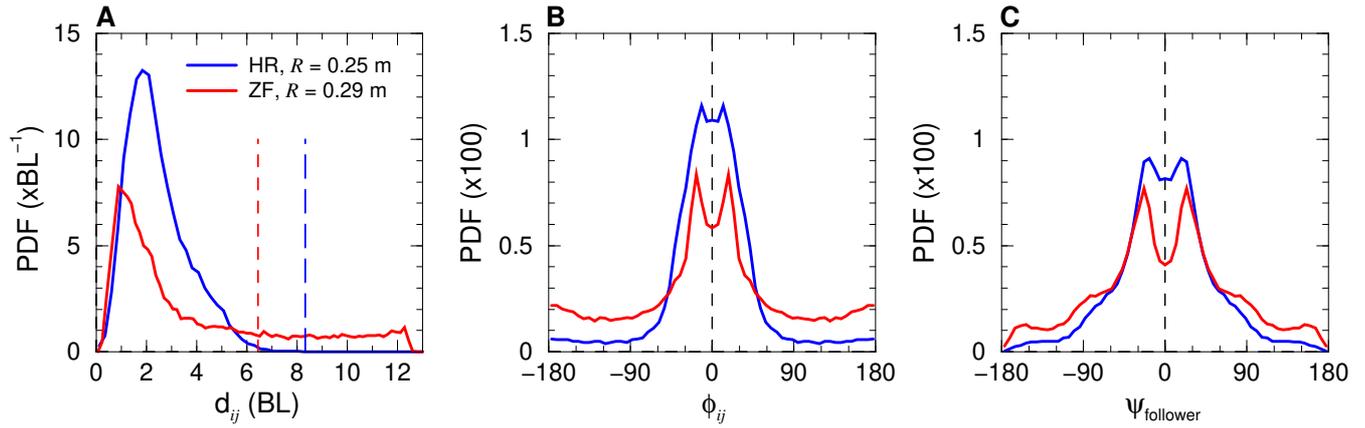


Figure 7: Probability density functions (PDF) of (A) distance between fish  $d_{ij}$ , (B) difference of heading  $\phi_{ij}$ , and (C) angle  $\psi_{\text{follower}}$  with which a fish  $i$  perceives its neighbour  $j$  when  $|\psi_{ij}| < |\psi_{ji}|$  (the “geometrical follower” is hence the fish which would have to turn the less to face the other fish, the latter being called the “geometrical leader”). Blue lines: *H. rhodostomus* in a circular arena of radius 0.25 m, red lines: *D. rerio* in a circular arena of radius 0.29 m. The vertical dashed lines in panel (A) denote the relative radius  $R/BL$  of the arena with respect to fish body length of each species, 8.3 in *H. rhodostomus* and 6.4 in *D. rerio* (with 1 BL = 0.03 m in *H. rhodostomus* and 0.045 m in *D. rerio*).

Derive equation from model:

$$\delta\phi(t) = \delta\phi_{\text{Att}}(d, \psi, \phi, v) + \delta\phi_{\text{Ali}}(d, \psi, \phi, v)$$

Assume separability of state variables:

$$\delta\phi_{\text{Att}}(d, \psi, \phi, v) = f(d)g(\psi)h(\phi)l(v)$$

$$\delta\phi_{\text{Ali}}(d, \psi, \phi, v) = \hat{f}(d)\hat{g}(\psi)\hat{h}(\phi)\hat{l}(v)$$

Discretize equation, calculate mean heading change in each box, put unknowns in the LHS:

$$\underbrace{f_i g_j h_k l_m}_{\text{unknowns}} + \underbrace{\hat{f}_i \hat{g}_j \hat{h}_k \hat{l}_m}_{\text{unknowns}} = \underbrace{\delta\phi_{ijkm}}_{\text{data}}$$

Write error function of overdetermined system:

$$\Delta(\vec{X}) = \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K \sum_{m=1}^M \epsilon_{ijkm} (\underbrace{\delta\phi_{S,ijkm}}_{\text{data}} - \underbrace{\delta\phi_{ijkm}}_{\text{unknowns}})^2$$

Minimize error function by finding the roots of its gradient with an iterative process:

Gradient: partial derivative with respect to  $f_i$ :

$$\frac{\partial \Delta}{\partial f_i} = 2 \sum_{j=1}^J \sum_{k=1}^K \sum_{m=1}^M \epsilon_{ijkm} g_j h_k l_m (\delta\phi_{S,ijkm} - \delta\phi_{ijkm})$$

Roots:

$$f_i = - \frac{\sum_{j=1}^J \sum_{k=1}^K \sum_{m=1}^M \epsilon_{ijkm} g_j h_k l_m (\hat{f}_i \hat{g}_j \hat{h}_k \hat{l}_m - \delta\phi_{ijkm})}{\sum_{j=1}^J \sum_{k=1}^K \sum_{m=1}^M \epsilon_{ijkm} (g_j h_k l_m)^2}$$

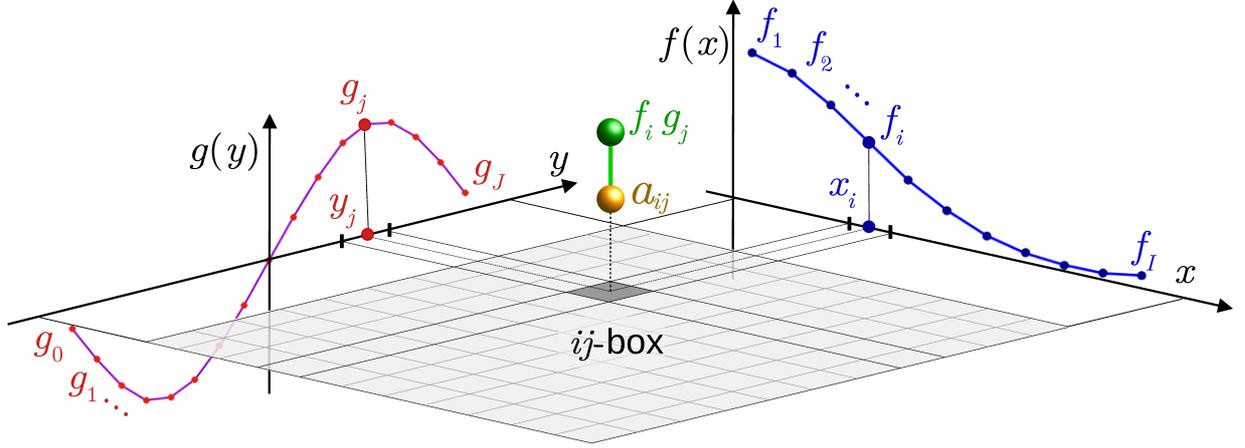
Explicit relations:

$$\begin{aligned} f_1 &= R_1(g_1, \dots, g_J, h_1, \dots, \hat{h}_K, \hat{l}_1, \dots, \hat{l}_M), \\ f_2 &= R_2(g_1, \dots, g_J, h_1, \dots, \hat{h}_K, \hat{l}_1, \dots, \hat{l}_M), \\ &\vdots \\ g_1 &= R_{I+1}(f_1, \dots, f_I, h_1, \dots, \hat{h}_K, \hat{l}_1, \dots, \hat{l}_M) \\ &\vdots \end{aligned}$$

Iterative process:  $\vec{X}^1 \rightarrow \vec{X}^2 \rightarrow \vec{X}^3 \rightarrow \dots$

initial guess  $\begin{bmatrix} f_i^0 \\ g_j^0 \end{bmatrix} \rightarrow \begin{aligned} f_i^{n+1} &= \lambda f_i^n + (1-\lambda)\tilde{f}_i^n \\ g_j^{n+1} &= \lambda g_j^n + (1-\lambda)\tilde{g}_j^n \end{aligned} \rightarrow \begin{bmatrix} f_i^* \\ g_j^* \end{bmatrix}$  fixed point

Figure 8: Detailed description of the successive substeps involved in the extraction of the interaction functions (step 4 of our method). For each substep, a short text describes what is done and the main formulas used to carrying it out.



$$\Delta(\vec{X}) = \sum_{i=1}^I \sum_{j=1}^J \epsilon_{ij} (f_i g_j - a_{ij})^2$$

Figure 9: Construction of the error function  $\Delta(\vec{X})$  in a simple two-dimensional case where  $a(x, y) = f(x)g(y)$ . After discretising the variables  $x$  and  $y$  in  $I$  and  $J$  intervals respectively, the values of  $a(x, y)$  are distributed in the resulting  $ij$ -boxes. The number of data in each box  $\epsilon_{ij}$  is counted, and the mean value of  $a(x, y)$  in each box,  $a_{ij}$ , is calculated. The square error is formulated in each box,  $(f_i g_j - a_{ij})^2$ , and summed up along all boxes. Starting from an initial guess of  $I + J$  values  $f_i^0, g_j^0, i = 1, \dots, I, j = 1, \dots, J$ , which produces a set of  $I \times J$  green balls (one per  $ij$ -box), the iterative process is carried out. At each iteration, the green balls of each box converge *globally* towards the yellow balls, until a minimum of the error  $\Delta(\vec{X})$  is reached. The final values  $f_i^n, g_j^n$  (blue and red points respectively) are then fitted by means of analytical expressions, which are the interaction functions represented by the blue and red solid curves. Note that  $n$  is the index of iteration, not the time discretisation index.

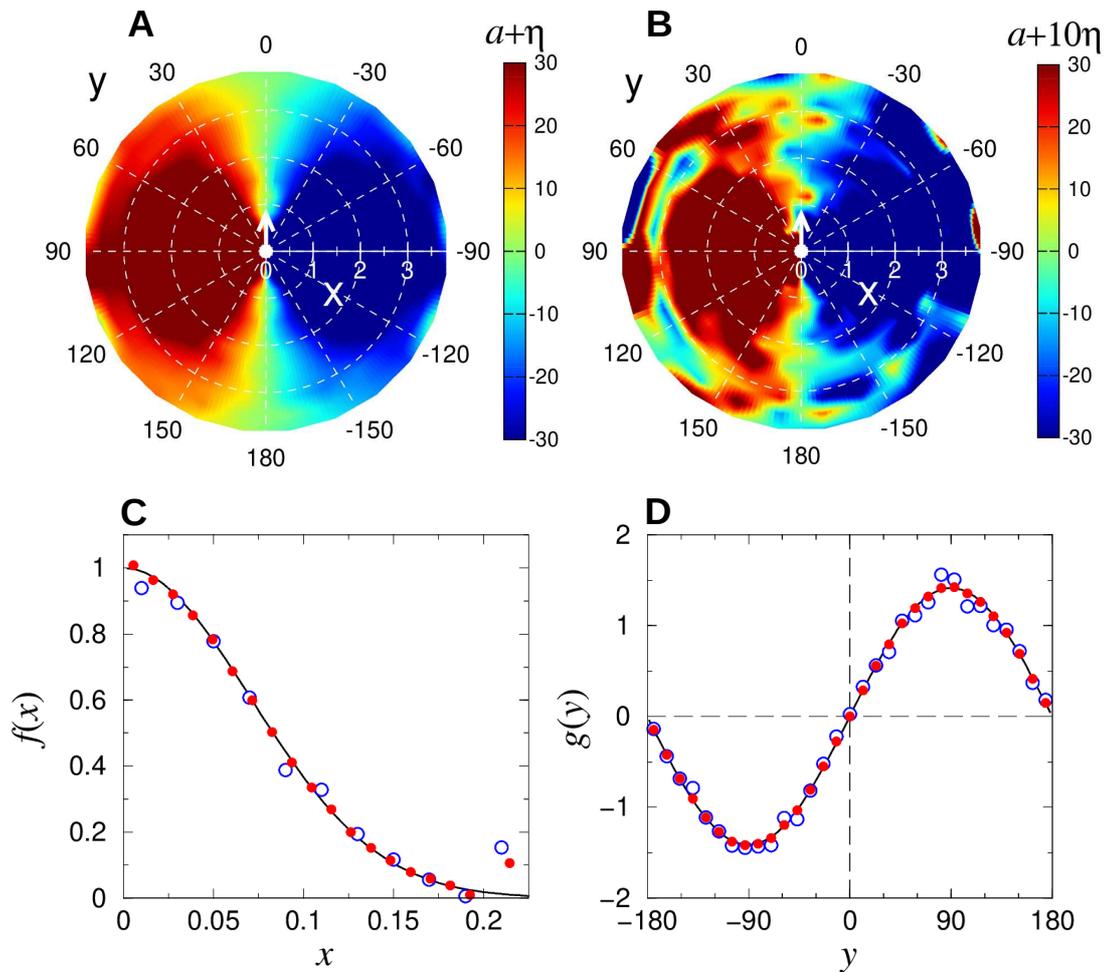


Figure 10: Force maps and reconstruction of the interaction functions in a simple case. (AB) Force maps of the function  $a(x, y) + \eta(x, y)$  for two levels of noise, of zero mean and amplitude (A) 0.35, which is the one observed in the experiments with *H. rhodostomus*, and (B) 3.5, a level of noise ten times higher. We used the same distribution of points  $\epsilon_{ij}$  than the one observed in the experiments with *H. rhodostomus*. (CD) Reconstruction of the interaction functions (C)  $f(x)$  and (D)  $g(y)$  extracted with our procedure (red dots and blue circles), compared to the analytical expressions  $\exp[-(x/x_0)^2]$  and  $\sqrt{2} \sin(y)$  respectively (black lines). The red dots correspond to the case of a noise of amplitude 0.35, where we used 20 nodes in  $x$  and 31 in  $y$ , and the blue circles to the case with a much higher level of noise (3.5) and with less nodes in  $x$ , 11 instead of 20. Note that the function  $g(y)$  is normalized.

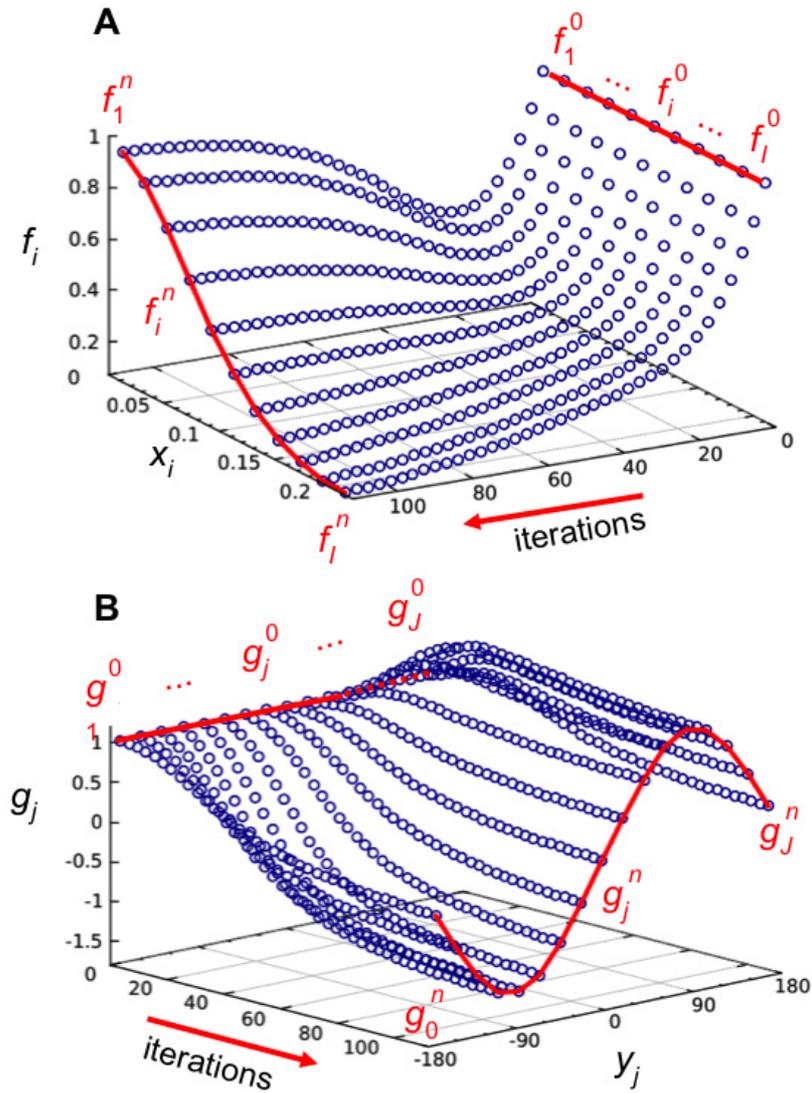


Figure 11: Evolution of the iterative process for the minimization of the error function  $\Delta(\vec{X})$  in the simple case where  $a(x, y) = \sqrt{2} \exp[-(x/x_0)^2] \sin(y)$ . Blue circles denote the successive values of the  $I + J$  components of the unknown functions  $f_i^n$  and  $g_j^n$  during the first 120 steps of the iterative process. Red lines show the initial guess of each function  $f_i^0 = g_j^0 = 1$ ,  $i = 1, \dots, I$  and  $j = 1, \dots, J$ , and the final state of convergence. Note that  $n$  is the index of iteration, not the time discretisation index.

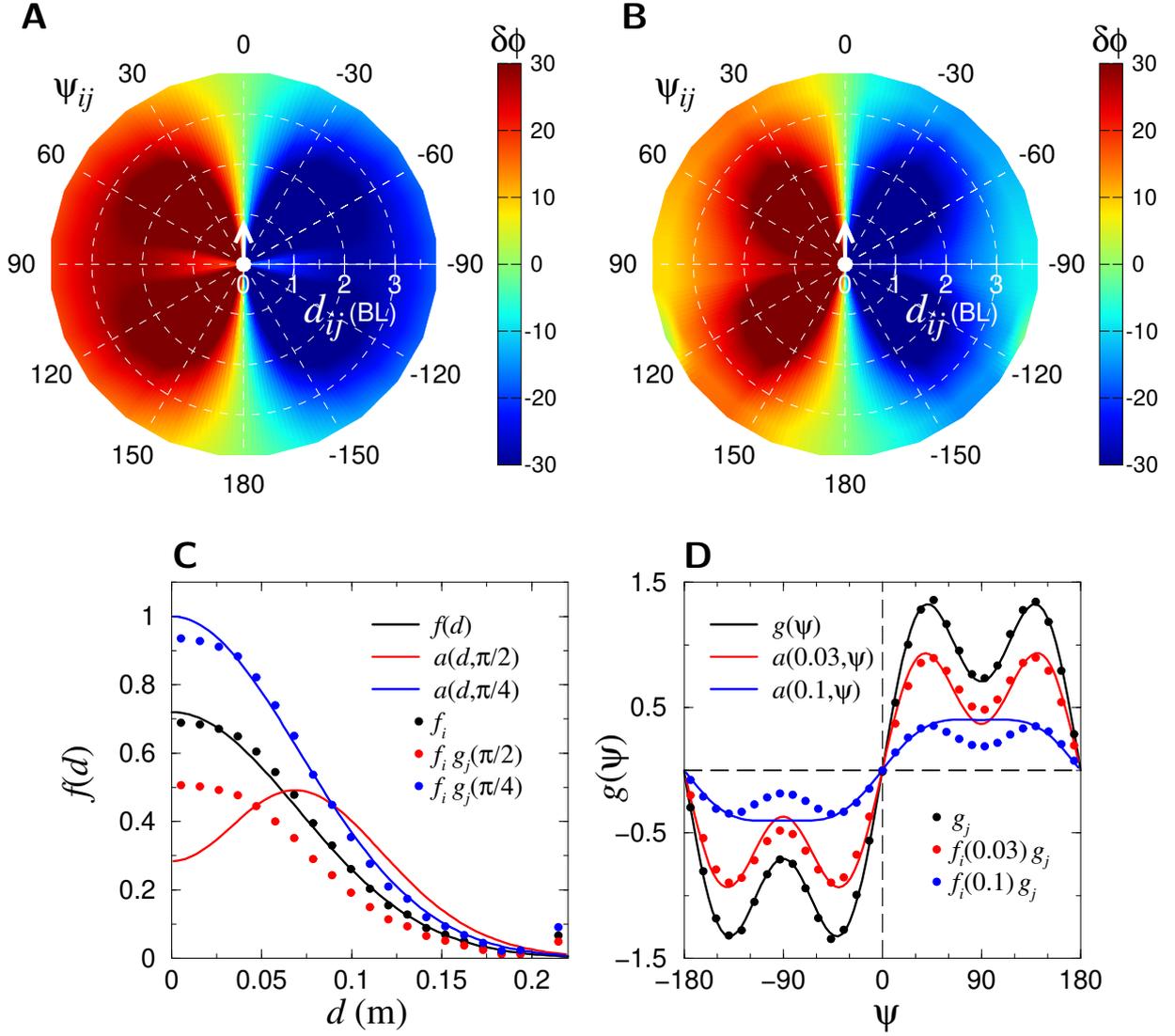


Figure 12: Colour maps of (A) the non-separable function  $b(d, \psi) = \sqrt{2}e^{-d^2/(2l_0^2)} \sin \psi + 1.13e^{-d^2/(2l_2^2)} \sin \psi \cos(2\psi)$  and (B) the solution found by the reconstruction procedure  $f(d)g(\psi) = 1.34e^{-d^2/(2l_0^2)} \sin \psi + 0.83e^{-d^2/(2l_2^2)} \sin \psi \cos(2\psi)$ . Red and blue solid lines: Cuts of the surface of  $b_1(d, \psi)$  for a fixed value of (C)  $\psi = \pi/2$  (red) and  $\psi = \pi/4$  (blue), and (D)  $d = 0.03$  (red) and  $d = 0.1$  (blue). Black dots denote the reconstructed functions  $f_i$  in (C) and  $g_j$  in (D). Solid black lines in both panels are the analytical expressions found for the reconstructed functions. Red and blue dots are cuts of the reconstructed solution  $f_i g_j$  corresponding to the respective cut of the non-separable function  $b_1(d, \psi)$ . Values of the cuts are  $f(0.03) = 0.67$ ,  $f(0.1) = 0.26$ ,  $g(\pi/2) = 0.73$  and  $g(\pi/4) = 1.36$ . A noise of the same amplitude as the one measured in the experimental data of *H. rhodostomus* was added to  $b_1(d, \psi)$  to create the data set for the reconstruction procedure. The same data structure (number of data per box) as the one found in real fish has been used. We have used  $I = 21$  and  $J = 31$ .

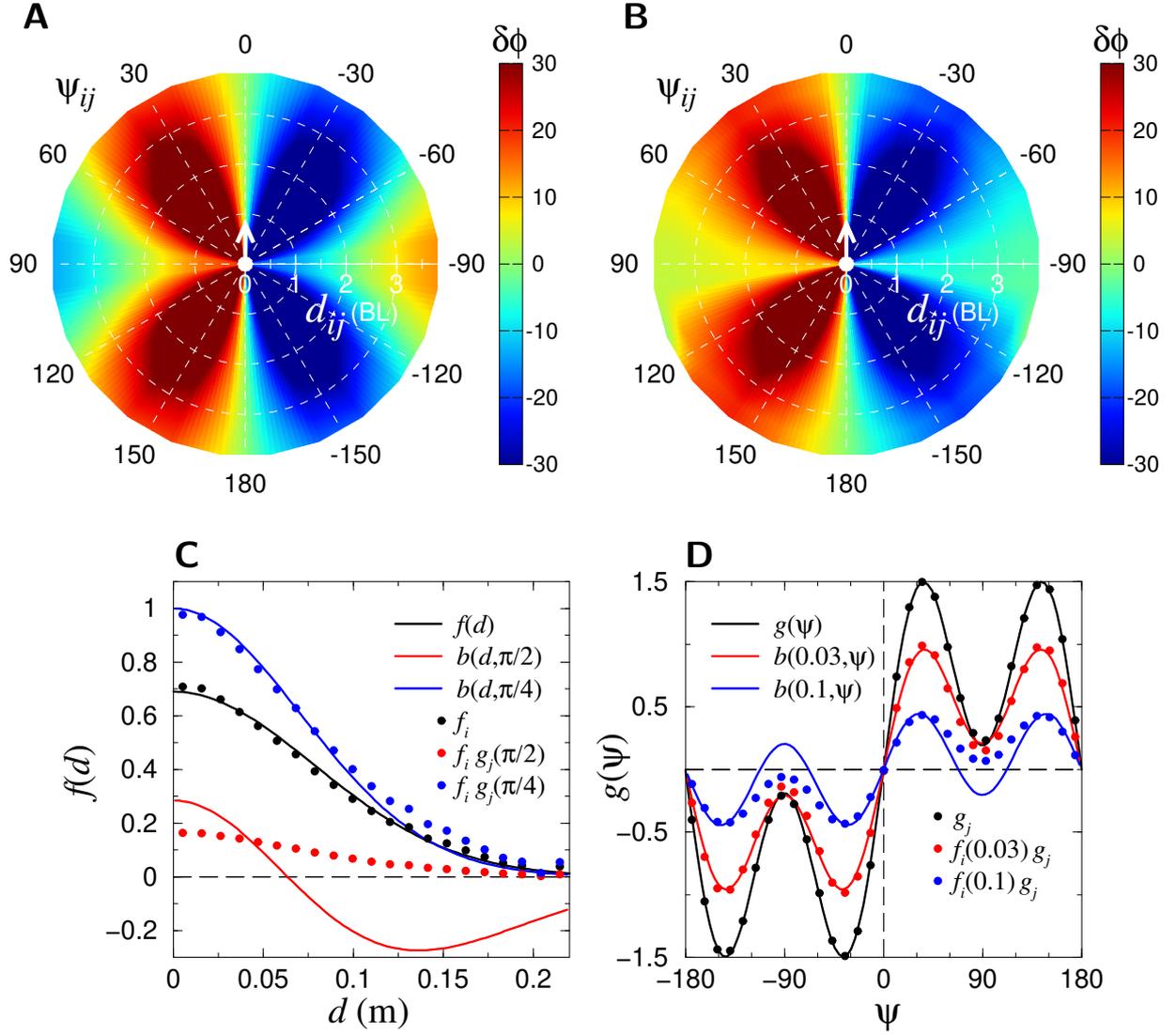


Figure 13: Colour maps of (A) the non-separable function  $b(d, \psi) = \sqrt{2}e^{-d^2/(2l_0^2)} \sin \psi + 1.13e^{-d^2/(2l_2^2)} \sin \psi \cos(2\psi)$  and (B) the solution found by the reconstruction procedure  $f(d)g(\psi) = 1.34e^{-d^2/(2l_0^2)} \sin \psi + 0.83e^{-d^2/(2l_2^2)} \sin \psi \cos(2\psi)$ . Red and blue solid lines: Cuts of the surface of  $b_2(d, \psi)$  for a fixed value of (C)  $\psi = \pi/2$  (red) and  $\psi = \pi/4$  (blue), and (D)  $d = 0.03$  (red) and  $d = 0.1$  (blue). Black dots denote the reconstructed functions  $f_i$  in (C) and  $g_j$  in (D). Solid black lines in both panels are the analytical expressions found for the reconstructed functions. Red and blue dots are cuts of the reconstructed solution  $f_i g_j$  corresponding to the respective cut of the non-separable function  $b_2(d, \psi)$ . Values of the cuts are  $f(0.03) = 0.23$ ,  $f(0.1) = 0.29$ ,  $g(\pi/2) = 0.66$  and  $g(\pi/4) = 1.38$ . A noise of the same amplitude as the one measured in the experimental data of *H. rhodostomus* was added to  $b_2(d, \psi)$  to create the data set for the reconstruction procedure. The same data structure (number of data per box) as the one found in real fish has been used. We have used  $I = 21$  and  $J = 31$ .

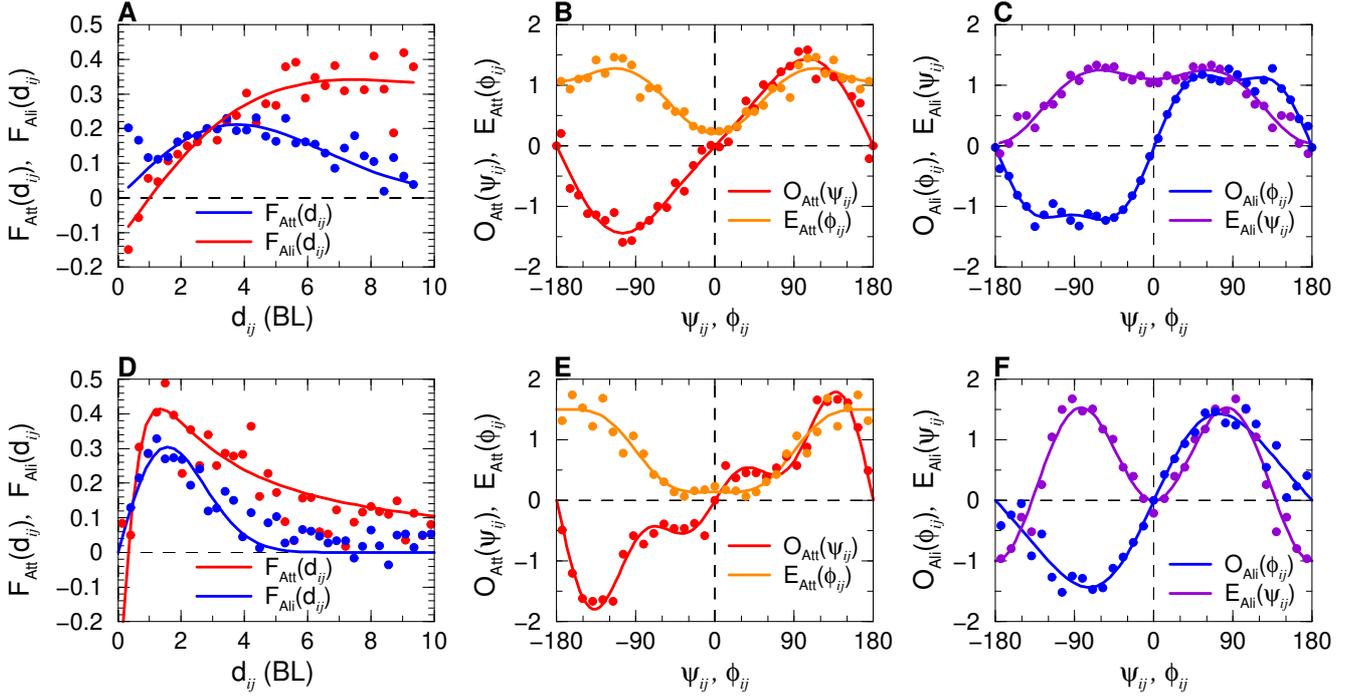


Figure 14: Analytical expressions of the social interaction functions of *H. rhodostomus* (ABC) and *D. rerio* (DEF) swimming in arenas of radius 0.25 and 0.29m respectively (solid lines), interpolating the discrete values extracted from the experimental data with the procedure described in Sec. 3 (dots). (AD) Intensity of the attraction (red) and alignment (blue), modulated by the angular functions of (BE) attraction and (CF) alignment.

### Box 1. Minimizing the error between data and a product form

Let us consider a  $N \times M$  two-dimensional array  $a_{ij}$ , for instance, built after averaging some quantity obtained from experiments on a  $N \times M$  grid. We look for the best approximation of  $a_{ij}$  under a product or separable form,  $a_{ij} \approx x_i \times y_j$ ,  $i = 1, \dots, N$ ,  $j = 1, \dots, M$  (see the system (7), in the more general case when more than two dimensions are involved). Note that under the product form, the number of unknown variables is only  $N + M$ , compared to the much larger size  $N \times M$  of the original array of data  $a_{ij}$ . In order to quantify the accuracy of the product description, we define the quadratic error function  $\Delta$ :

$$\Delta(x_1, \dots, x_N, y_1, \dots, y_M) = \sum_{i=1}^N \sum_{j=1}^M \epsilon_{ij} (x_i y_j - a_{ij})^2. \quad (27)$$

Here  $\epsilon_{ij}$  is the number of data that fall in the cell  $ij$ . For systems that derive from physical and biological phenomena, it is fundamental to modulate the contribution of each cell, not only to give a higher weight to the more frequent values, but also to preserve actual correlations arising between the variables.

The goal is thus to find a set of values  $x_1, \dots, x_N, y_1, \dots, y_M$  that minimizes the error  $\Delta$ .  $\Delta = 0$  implies that  $x_i y_j = a_{ij}$ , for all  $i$  and  $j$ , and that the original data had exactly a product form. Finding the minima of a function requires finding the zeros of its derivative, which, in high dimensions, is the *gradient vector*  $\vec{\nabla} \Delta$ , given by the partial derivatives of  $\Delta$  with respect to its components; see the Box *Finding minima* below.

#### Finding minima

A minimum of a one-variable function  $f(x)$  is a point  $x_m$  where the derivative vanishes,  $f'(x_m) = 0$ , and the second derivative is positive,  $f''(x_m) > 0$ .

In several dimensions, the “derivative” of a function  $F : \mathbb{R}^D \rightarrow \mathbb{R}$  is the *gradient vector*  $\vec{\nabla} F(\vec{x})$ , whose components are given by the partial derivatives of the function with respect to each component of  $\vec{x}$ :  $\vec{\nabla} F(\vec{x}) = (\partial F / \partial x_1, \partial F / \partial x_2, \dots, \partial F / \partial x_D)$ . The gradient vector points in the direction of maximum variation of  $F$ , and is therefore zero (*i.e.*, equal to the null vector  $\vec{0}$ ) when  $\vec{x}$  is a extremum, and in particular, a minimum:

$$\partial F / \partial x_1 = 0, \quad \partial F / \partial x_2 = 0, \dots, \quad \partial F / \partial x_D = 0. \quad (28)$$

In principle, to ensure that the extremum is indeed a minimum, one also has to check that the Hessian matrix of second-order partial derivatives has only strictly positive eigenvalues, a condition generically satisfied for squared error functions like the one considered here.

Differentiating (27) with respect to  $x_k$  gives

$$\frac{\partial \Delta}{\partial x_k} = \sum_{i=1}^N \sum_{j=1}^M \frac{\partial}{\partial x_k} [\epsilon_{ij} (x_i y_j - a_{ij})^2] = 2 \sum_{j=1}^M \epsilon_{ij} y_j (x_k y_j - a_{kj}), \quad (29)$$

where we used the fact that  $x_i y_j$  does not depend on  $x_k$ , if  $i \neq k$ . Then,

$$\frac{\partial \Delta}{\partial x_k} = 2x_k \sum_{j=1}^M \epsilon_{ij} y_j^2 - 2 \sum_{j=1}^M \epsilon_{ij} y_j a_{kj}, \quad (30)$$

and the conditions  $\partial \Delta / \partial x_i = 0$  and  $\partial \Delta / \partial y_j = 0$  for all  $i$  and  $j$  can be rewritten as

$$x_i = \frac{\sum_{j=1}^M \epsilon_{ij} y_j a_{ij}}{\sum_{j=1}^M \epsilon_{ij} y_j^2}, \quad i = 1, \dots, N, \quad y_j = \frac{\sum_{i=1}^N \epsilon_{ij} x_i a_{ij}}{\sum_{i=1}^N \epsilon_{ij} x_i^2}, \quad j = 1, \dots, M. \quad (31)$$

This is a system of  $N + M$  equations and unknowns that can be solved with different methods. Due to the large dimension of the systems arising in social interaction analysis, iterative methods are often used. See Box 2.

## Box 2. Fixed point iterations

A point  $x^* \in \mathbb{R}$  is a fixed point of a function  $f : \mathbb{R} \rightarrow \mathbb{R}$  if  $f(x^*) = x^*$ . Fixed points can be found under certain conditions<sup>a</sup> by means of the iterative method

$$x_{n+1} = f(x_n). \quad (32)$$

Starting from an initial point  $x_0$ , the method builds a sequence  $x_1, x_2, x_3, \dots$  that converges to (one of) the fixed point(s) of  $f$ ; see Fig. 15.

In  $d$  dimensions, a vector  $\vec{x}^* \in \mathbb{R}^d$  is a fixed point of a function  $\vec{F} : \mathbb{R}^d \rightarrow \mathbb{R}^d$  if  $\vec{F}(\vec{x}^*) = \vec{x}^*$ .

When the dimension of the system  $d = N + M$  is very large, and in order to improve the stability of the recursion dynamics, it is convenient to use a relaxation method,

$$\vec{x}_{n+1} = \lambda \vec{x}_n + (1 - \lambda) \vec{F}(\vec{x}_n), \quad (33)$$

where  $\lambda \in (0, 1)$  is the weight of the previous iteration

<sup>a</sup>The function  $f$  must be contractive and the initial value of the iterations must be sufficiently close to the fixed point.

in the value of the new iteration, averaged with what would have been the new iteration.

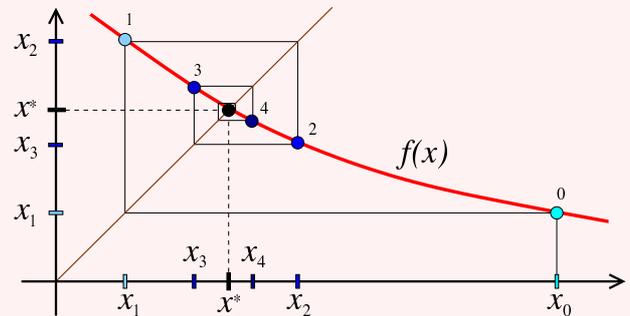


Figure 15: Iterations converging to  $(x^*, x^*)$ . Red line: function  $f(x)$ ; brown line:  $y = x$ ; thin polygonal: iteration process; coloured dots: successive values of  $x_n$ .