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

R. Escobedo¹, V. Lecheval³, V. Papaspyros⁴, F. Bonnet⁴, F. Mondada⁴, C. Sire² and G. Theraulaz¹,⁵

¹Centre de Recherches sur la Cognition Animale, Centre de Biologie Intégrative (CBI), and ²Laboratoire de Physique Théorique, Centre National de la Recherche Scientifique (CNRS) & Université de Toulouse – Paul Sabatier, 31062 Toulouse, France
³Department of Biology, University of York, York Y010 5DD, UK
⁴MOBOTS group, Bioinformatics laboratory, École Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland
⁵Centre for Ecological Sciences, Indian Institute of Science, Bengaluru, India

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 datasets, 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.

This article is part of the theme issue ‘Multi-scale analysis and modelling of collective migration in biological systems’.

1. Introduction

The identification and characterization of interactions between the constituent elements of a living system are 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 cognitive 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 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 [7–11]. 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 are 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 of 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 associating these measures with the relative state of the individuals involved in these interactions [12]. The relative state of an individual \( i \) with respect to a focal individual \( j \) is determined by the distance \( d_{ij} \) between \( i \) and \( j \), its relative velocity \( \psi_{ij} \), its angular position with respect to \( i \), \( \psi_{ij} \) (viewing angle), and its relative heading \( \phi_{ij} \) (figure 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. Thus, the task is to relate the heading variation of a focal fish \( \delta \phi \), with its state variables, that is, to find a function \( \delta \phi(d, \psi, \phi) \).

In this article, we first show that force maps, which 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 visualize the strength and direction of behavioural changes. However, force maps have strong limitations that can induce profound misunderstandings.

(a) Visualization 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 dataset is a sequence of \( N \) triplets \((x_n, y_n, f_n)\), where the index \( n \) denotes for example the instant of time \( t_n \), \( n = 1, \ldots, N \). To build a force map of this function, the \((x, y)\)-space is discretized in \( I \times J \) boxes of the form...
Figures 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, in the presence of a conspecific in a circular tank. In the case of *Hemigymmus rhodostomus* (figure 2), we used a tank of radius $R = 0.25\, \text{m}$, about 8.3 times the body length (BL) of the fish, and in the case of *Danio rerio* (figure 3), a tank of radius $R_x = 0.29\, \text{m}$, about 6.4 BL. Both figures 2a and 3a 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 figures 2b and 3b 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*, figure 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 $\leq 30^\circ$, probably owing 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 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 BL, and repulsed when its neighbour is too close ($d_{ij} < 1$ BL). The force map in figure 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

\[ [\hat{x}_i, \hat{x}_{i+1}] \times [\hat{y}_j, \hat{y}_{j+1}], \] where the nodes $\hat{x}_i$ and $\hat{y}_j$ are given by

\[ \hat{x}_i = \hat{x}_{\min} + (i - 1)(\hat{x}_{\max} - \hat{x}_{\min})/I, \quad i = 1, \ldots, I + 1 \]

and

\[ \hat{y}_j = \hat{y}_{\min} + (j - 1)(\hat{y}_{\max} - \hat{y}_{\min})/J, \quad j = 1, \ldots, J + 1, \]

and the data $(x^i, y^j, f^o)$ is placed in the $ij$-box such that $x^i$ is in $[\hat{x}_i, \hat{x}_{i+1}]$ and $y^j$ is in $[\hat{y}_j, \hat{y}_{j+1}]$. Then, the number of data $e_{ij}$ in the $ij$-box and the mean value $f_{ij}$ of the values of $f^o$ that fall 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, $x_i = (\hat{x}_i + \hat{x}_{i+1})/2$, $y_j = (\hat{y}_j + \hat{y}_{j+1})/2$. 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].

**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 $d_{ij}$ and $\phi_{ij}$, and (d) the relative heading $\phi_{ij}$ and the front–back distance $d_{ij}$. The colour scale shown on the right represents the average value of heading change $\delta\phi$. Distances are measured in body lengths (BL), angles in degrees.
small. This force map reveals that a fish tends to align with its neighbour. In *D. rerio*, figure 3a shows that the focal fish turns towards its neighbour when they are close to each other \( d_{ij} \approx 1–3 \text{ BL} \) or when the neighbour is located behind the focal fish \( (|\psi_{ij}| > 90^\circ, \text{whatever the distance}) \). 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 the neighbour’s heading is less intense than in *H. rhodostomus*, as shown by the wide white regions in figure 3b. When fish are beyond 3 BL from each other, the focal fish turns to adopt the orientation of its neighbour. At short distances, the behaviour is more complex, with changes of smaller size than those observed in *H. rhodostomus*.

The visualization 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, which leads a fish to turn so as to adopt the same heading as its neighbour. However, nothing can be deduced from these colour maps about what happens when the two 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 \( (d_{ij}, \psi_{ij}, \phi_{ij}) \) 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 into the left-right (LR) and front-back (FB) distances [13]. In figure 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 figure 2a that are in the left semicircle of radius 1 BL are averaged in a single point where \( d_{ij}^{LR} = -1 \text{ BL} \) (where \( d_{ij}^{LR} \) is the LR distance of fish \( j \) with respect to fish \( i \) because these points are at 1 BL to the left of the focal fish. Then, the third variable \( \phi_{ij} \) is used to expand this averaged point on a vertical line with different values of \( \phi_{ij} \) in a system of reference with coordinates \( (d_{ij}^{LR}, \phi_{ij}) \), giving rise to figure 2c. Similarly, the upper semicircles of figure 2a are averaged on the \( d_{ij}^{FB} \)-axis in figure 2d.

**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 on 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) *Hemigrammus rhodostomus* in an arena of radius 0.25 m, and (b) *D. rerio* in an arena of radius 0.29 m, when the focal fish is at least at 2 BL (6 cm) in *H. rhodostomus* and 1 BL (4.5 cm) 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 \( i \).

Force maps in figures 2c,d and 3c,d provide additional information about individual fish behaviour. In *H. rhodostomus*, figure 2c 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 \( |d_{ij}^{LR}| > 2 \text{BL} \) and \( \phi_j \approx \pm 90^\circ \)). In the white horizontal region, the focal fish maintains its heading when it is aligned with its neighbour (\( |\phi_j| < 10^\circ \)), whatever the horizontal distance between them. Figure 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 \( d_{ij}^{FB} < -2 \text{BL} \)). In *D. rerio*, the colour of each vertical half-plane of figure 3c is almost uniform, except for some regions close to the focal fish \( (d_{ij}^{LR} \approx \pm 1 \text{BL}) \) at \( \phi_j \approx \pm 90^\circ \), and some distant regions located at \( d_{ij}^{LR} \approx \pm 3 \text{BL} \) and \( \phi_j \approx \pm 135^\circ \). 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 figure 3d, one can see that the fish turns to adopt the same direction as that of its neighbour when the neighbour 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 \( d_{ij}^{FB} > 2 \text{BL} \)) or behind it and not very close to it (regions where \( d_{ij}^{FB} < -1 \text{BL} \)). This is a different and more complex behaviour than the one observed in *H. rhodostomus*, where heading changes depend on \( \phi_j \) but not on \( d_{ij}^{FB} \). In *D. rerio*, we observe the opposite, and the dependence on the FB distance is more complex than in *H. rhodostomus*. However, neither \( d_{ij}^{LR} \) nor \( d_{ij}^{FB} \) separately, can determine the distance at which the neighbour is (e.g. a fish \( j \) located at 1 BL to the left can be at 0.5 or 2 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 the contribution of intermediate effects such as attraction and alignment to be disentangled.

(b) 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 to force maps. For example, data can be scarce; not only because they are difficult to collect but also because the phenomenon under observation rarely produces 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, as well as 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, figure 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, e.g., information from the frontal region in figure 4a is more relevant to describe the behaviour of *H. rhodostomus* than the information from the same region in figure 4b is to describe the behaviour of *D. rerio*.

Thus, the relative simplicity to visualize 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 two, at most three. 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.
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 BL 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 is the case of the heading change \(\delta \phi\) with respect to the angle of perception \(\psi_i\) and the relative heading \(\phi_i\) in *H. rhodostomus*, where \(\delta \phi(d_{ij}, \psi_i, \phi_i) = -\delta \phi(d_{ij}, -\psi_i, \phi_i) = -\delta \phi(d_{ij}, \psi_i, -\phi_i)\) [12]. Then, for instance, huge variations of \(\delta \phi\) with respect to \(\psi_i\) but of different sign, can cancel each other and yield a small average value, as if \(\delta \phi\) was almost independent of \(\psi_i\). 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. This would mean for instance that the probability of a fish \(i\) being in the state \(d_{ij} = 1\) BL, \(\psi_j = 90^\circ\) and \(\phi_i = 10^\circ\) with respect to a focal fish \(j\) is the same as the probability of being at \((d_{ij}, \psi_j, \phi_j) = (2, 0, 180)\), 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 owing to the strong correlations between the states variables along the actual trajectories. For instance, averaging the interaction between two 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 \(\delta \phi(\psi, \phi) = g(\psi)h(\phi)\int_0^\infty C(r, \psi, \phi) dr\). 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 two fish are close, they are likely to be aligned (\(\phi \approx 0\)), whereas if they are far apart, 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, such as, e.g. the relative velocity \(v_{ij}\) or the interaction of fish with an obstacle (such as the wall of a tank), which would require two more variables \(r_{iv}\) and \(\theta_{iw}\), the distance and angle to the wall, respectively. Even for the depicted intermediate values, the precise contributions of the different variables to the heading change are still entangled. For instance, how does \(\phi_i\) vary when \(j\) points to the left 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 by 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 for the heading change \(\delta \phi\), which depends on the state variables \(d_{ij}, \psi_j\) and \(\phi_j\) 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 identification of 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. Figures 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 BL of individuals. Are the differences observed in the previous maps of each species due to the experimental 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 BL, 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 essentially consists in defining this framework
and deriving the mathematical model describing the relation
between the quantities used to quantify the behaviour 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 calculation
of the average in each box of the discretized 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.

(a) Outline of the method

Figure 5 provides a general overview of the method used to
design the mathematical model and to extract the social inter-
action 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 figure 5, as the model
predictions allow better targeted experiments to be performed,
which in turn allow the model to be refined: the experiments
feed the model, and in turn the model helps the design of experi-
ments. Figure 6 shows the same five steps in a detailed flow chart
that includes the test of the model.

(i) 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 $u(t)$ resulting
from a tracking procedure conveniently purified from identifi-
cation errors (step 1 in figures 5 and 6). The velocity vector $v_i$
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.

(ii) 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 figures 5 and 6).

In the case considered here, the individual state of a fish is
characterized by its position and velocity vectors, or, equival-
ently, 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 $\phi_i = \phi_j - \phi$. The heading angle change $\delta \phi_{ij}$
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 observ-
ables 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). Figure 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 exclu-
sively on the experimental data and is independent of a
model. Eventually, force maps can be drawn.

(iii) Modelling hypotheses

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

This process has been described in detail in [12] in the
case of $H$. rhodostomus; here we simply summarize it, as it is
identical for $D$. rerio. We first consider that fish move in a
straight line during short time intervals of different durations.
These time intervals are separated by instantaneous 'kicks'
during which fish adjust their direction and make an
abrupt acceleration, after which the speed decays exponen-
tially 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
well reproducing the experimental ones, as described in [12].

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

$$
\begin{align*}
\dot{u}(t + dt) &= u(t) + l(t) e(\phi(t) + dt), \\
\dot{\phi}(t + dt) &= \phi(t) + \delta \phi(t),
\end{align*}
$$

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

$$
\delta \phi(t) = \delta \phi_{ij}(t) + \delta \phi_{ij}(t) + \delta \phi_{ij}(t),
$$

(3.1)
therein dependence on relative orientation cycle starts again, which includes the realization of a new set of experiments or the design of a new set-up, to check the predictions of the model or to deepen our understanding of the mechanisms at work.

combined linearly (additive hypothesis).

heading variation of a fish, and that these contributions are the wall when the fish is close to a tank wall, and 

action functions determining the contributions of a

neighbouring fish and of the obstacles in the environment

Figure 5. General overview of the methodology. The successive steps carried out in our method are labelled 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 \( \phi_w \) and the relative orientation \( \phi_\nu \) and PDF of \( d_{ij} \); (3) model equations of the heading angle changes \( \delta \phi \); (4) extracted interaction functions \( F_{\text{Att}}, F_{\text{Ali}}, E_{\text{Att}}, E_{\text{Ali}}, O_{\text{Att}}, O_{\text{Ali}} \) and \( \delta \phi_{\nu} \) and their dependence on \( d_{ij}, \psi_y, \phi_y \); (5) trajectories resulting from the numerical simulations of the model. The darker blue arrow illustrates the possibility that the cycle starts again, which includes the realization of a new set of experiments or the design of a new set-up, to check the predictions of the model or to deepen our understanding of the mechanisms at work.

where \( \delta \phi_k \) is a random angle change accounting for the spontaneous decisions of the fish, \( \delta \phi_{\nu} \) 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

\[
\delta \phi(t) = \delta \phi_{\text{Att}}(t) + \delta \phi_{\text{Ali}}(t) + \delta \phi_{\nu}(t).
\]  

Equations (3.1) and (3.2) are based on the hypotheses that \( \delta \phi_{\nu} \), \( \delta \phi_{\nu}, \delta \phi_{\text{Ali}} \), and \( \delta \phi_{\text{Ali}} \) are the main contributions to the heading variation of a fish, and that these contributions are combined linearly (additive hypothesis).

(iv) Extraction of interaction 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 figure 8.

Discretizing the six 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, equation (3.1) can be rewritten as a system of \( \Omega \) equations and \( \Omega \) unknowns,

\[
\delta \phi_{\nu}(t, \psi_y, \phi_y, v_m) + \delta \phi_{\nu}(d_{ij}, \psi_y, \phi_y, v_m) = \delta \phi_{\nu}(t, \psi_y, \phi_y, v_m)
\]  

where the noise term \( \delta \phi_{\nu} \) vanishes as it is assumed to have zero mean in each box. We use the convention that unknowns are written in the left-hand side of the equation and the known values in the right-hand side. Even if one could build seven-dimensional representations, the contributions of each variable would still be entangled, and intermediate functions impossible to detect in the corresponding behavioural maps of social 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]:

\[ \delta \phi_w (r_w, \theta_w) = f(r_w) g(\theta_w), \]

\[ \delta \phi_{\Delta u} (d, \phi, \psi, v) = f(d) g(\phi) h(d) l(v) \]

and

\[ \delta \phi_{\Delta d} (d, \phi, \psi, v) = f(d) g(\phi) h(d) l(v). \]

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

For the sake of simplicity, we consider here the case where fish are far enough from the tank wall that the contribution of \( \delta \phi_w \) to the heading variation can be neglected with respect to the effects of social interactions. The procedure can easily be extended to extract both the social interactions and the interaction with the wall, but at the cost of more complicated notation. We refer the reader to the work of Calovi et al. [12], in which the disentangling of the combined effects of a tank wall and social interactions between fish is described in detail.

Hence, we are left with four state variables, so that the heading change \( \delta \theta (t) \) is averaged in four-dimensional \( ijm \)-boxes. We introduce the notation \( f(d) = f_i, g(\phi) = g_1, ..., l(\theta_u) = l_u \), so that, after removing the interaction with the wall, the system (3.3) becomes

\[ f_i g_j h_k l_m + \hat{f}_i \hat{g}_j \hat{h}_k \hat{l}_m = \delta \phi_{ijkm}. \]

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

The goal is now to solve this system of equations. This will provide the discrete values of the interaction functions (the unknowns \( f_i, g_j, ..., l_u \)) as a function of the known values \( \delta \phi_{ijkm} \) (calculated from the experimental data as the mean heading change in each box). Systems with more equations than unknowns are called overdetermined systems and rarely have a solution. One way of overcoming this problem consists in reducing the overdetermined system to a solvable one by minimizing the error \( \Delta \) with which the overdetermined system is satisfied by a candidate solution that is updated iteratively. This corresponds to step 4.1 in the flowchart figure 6, which is described in detail in box 1 and in §3b below for a simple 2D case where the equation is \( f_i g_j = a_i \). Figure 6 shows how the error function is built in this simple case.

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

\[ \Delta (X) = \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} \sum_{m=1}^{M} \epsilon_{ijkm} (\delta \phi_{ijkm} - \delta \phi_{ijkm})^2, \]

where \( X \) is the vector of dimension \( D \)

\[ X = (f_1, ..., f_I, g_1, ..., g_J, h_1, ..., h_K, l_1, ..., l_M), \]

\[ \hat{f}_i, \hat{f}_1, \hat{g}_j, ..., \hat{g}_J, \hat{h}_i, ..., \hat{h}_K, \hat{l}_m, \hat{l}_1, ..., \hat{l}_M, \]

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

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

---

**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. (Online version in colour.)
For the gradient to be zero, we must have

\[
\frac{\partial \Delta}{\partial f_i} = 0, \quad \frac{\partial \Delta}{\partial f_j} = 0, \quad \ldots, \quad \frac{\partial \Delta}{\partial f_t} = 0, \quad \ldots, \quad \frac{\partial \Delta}{\partial f_M} = 0.
\]

For the component \( f_i \), the partial derivative of \( \Delta \) is

\[
\frac{\partial \Delta}{\partial f_i} = 2 \sum_{j=1}^{K} \sum_{k=1}^{M} \epsilon_{jkm} g_{j} h_{i} l_{m} (\delta \phi_{jkm} - \delta \phi_{jkm}),
\]

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

\[
f_i = \frac{-\sum_{j=1}^{K} \sum_{k=1}^{M} \epsilon_{jkm} g_{j} h_{i} l_{m} (f_{j} g_{j} h_{i} l_{m} - \delta \phi_{jkm})}{\sum_{j=1}^{K} \sum_{k=1}^{M} \sum_{m=1}^{M} \epsilon_{jkm} (g_{j} h_{i} l_{m})^2}.
\]

The corresponding equations for the other components are derived in the same way and all have the same structure where each component is given by an explicit combination of the other components and the known values \( \delta \phi_{jkm} \) and \( \epsilon_{jkm} \). For instance, the equation of \( g_i \) is obtained by replacing \( g_j \) by \( f_i \), \( j = 1 \) by \( i = 1 \), and \( j \) by \( l \), and the equation for a function with a hat (e.g. \( \hat{f}_i \)) is obtained by replacing the functions that have a hat by the same functions without the hat, and vice versa (i.e. \( \hat{g}_i \) is replaced by \( g_i \) and so on).

The resulting set of explicit equations for each component can be written in the following form,

\[
f_1 = R_1(g_1, \ldots, g_i, h_1, \ldots, \hat{h}_k, \hat{l}_1, \ldots, \hat{l}_M),
\]

\[
f_2 = R_2(g_1, \ldots, g_i, h_1, \ldots, \hat{h}_k, \hat{l}_1, \ldots, \hat{l}_M),
\]

\[
\vdots
\]

\[
g_1 = R_1(f_1, \ldots, f_i, h_1, \ldots, \hat{h}_k, \hat{l}_1, \ldots, \hat{l}_M),
\]

\[
\vdots
\]

\[
\hat{l}_M = \hat{R}_D(f_1, \ldots, f_i, g_1, \ldots, \hat{g}_1, h_1, \ldots, \hat{h}_k).
\]

To find the solution of such a system, it is possible to write it in a compact form as \( X = R(X) \), where \( R \) is a function from \( \mathbb{R}^{12} \) to \( \mathbb{R}^{12} \), and consider the problem as finding a fixed point \( X_0 \) of the function \( R \). This is done by means of an iterative method, as explained in detail in box 2, which also shows how this method works in the 1D case. Again, note that the number \( D = 2(I + J + K + M) \) of fitting parameters is much smaller than the number of boxes \( I \times J \times K \times M \).

In the end, the procedure provides the values of each interaction function at the points representing each box minimizing the error functions. Depending on the number of boxes, a satisfactory analytical form of each one-variable function can be obtained, based on physical principles and specific observations. This part corresponds to the step 4.3 in figure 6 and depends on the specific phenomenon under study; it is detailed in §4 for the case of fish swimming in pairs. For example, the function of the angle of perception \( g(\psi) \) must be odd, i.e. \( g(-\psi) = -g(\psi) \), because the attractive effect of a neighbour on the heading change of a focal fish has the same intensity wherever the neighbour is located at the right or at the left of the focal fish, but has opposite directions, and hence opposite signs, in each case: if the neighbour is at the right (resp. left) side, the focal fish would turn right (resp. left) to approach the neighbour. Physical properties of the interactions at play, such as the exponential decay of some interactions, must be taken into account and guide the choice of the final analytical expressions.

(v) Numerical simulations of the model

The last step consists in performing numerical simulations of the model with the double purpose of (1) verifying that the simulation results are in good agreement with the experimental results, and (2) making new predictions that can be ultimately confirmed by new experiments. If one of these two points is not satisfactorily verified, then it is necessary to go back to a previous step: step 3 to revise or reformulate the model, step 2 to use alternative state variables or observables, or even step 1 to carry out new experiments or measures of the data (figure 6).

(b) Application to a simple case study

In order to illustrate how the extraction procedure (step 4 of our methodology in figure 6) can be used, we have produced artificial data for a simple case in which the interaction
derive equation from model: 
\[ \delta \phi (t) = \delta \phi_{A11}(d, \psi, \phi, v) + \delta \phi_{A11}(d, \psi, \phi, v) \]

assume separability of state variables: 
\[ \delta \phi_{A11}(d, \psi, \phi, v) = f(d) g(\psi) h(\phi) l(v) \]
\[ \delta \phi_{A11}(d, \psi, \phi, v) = \hat{f}(d) \hat{g}(\psi) \hat{h}(\phi) \hat{l}(v) \]

discretize equation, calculate mean heading change in each box, put unknowns in the LHS: 
\[ f_i g_j h_k l_m + f_i g_j h_k l_m = \delta \phi_{ijkmn} \]
unlike unknowns data

write error function of overdetermined system: 
\[ \Delta(\vec{X}) = \sum_{i=1}^{J} \sum_{j=1}^{K} \sum_{k=1}^{M} \sum_{m=1}^{L} e_{ijk}(\delta \phi_{ijk} - \delta \phi_{ijkm})^2 \]
data unknowns

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

gradient: partial derivative with respect to \( f_i \):
\[ \frac{\partial \Delta}{\partial f_i} = 2 \sum_{j=1}^{J} \sum_{k=1}^{K} \sum_{m=1}^{L} e_{ijk}(\delta \phi_{ijk} - \delta \phi_{ijkm}) \]
roots:
\[ f_i = \frac{1}{\sum_{j=1}^{J} \sum_{k=1}^{K} \sum_{m=1}^{L} e_{ijk}(\delta \phi_{ijk} - \delta \phi_{ijkm})^2} \]
explicit relations:
\[ f_1 = R_1(g_1, \ldots, g_J, h_1, \ldots, h_K, l_1, \ldots, l_M), \]
\[ f_2 = R_2(g_1, \ldots, g_J, h_1, \ldots, h_K, l_1, \ldots, l_M), \]
\[ \vdots \]
\[ g_1 = R_1(f_1, \ldots, f_J, h_1, \ldots, h_K, l_1, \ldots, l_M), \]
\[ \vdots \]
iterative process: \[ \vec{X}^1 \rightarrow \vec{X}^2 \rightarrow \vec{X}^3 \rightarrow \ldots \]
initial \[ f_i^0 \rightarrow f_i^{n+1} = \lambda f_i^n + (1-\lambda) g_i^n \rightarrow f_i \] fixed
guess \[ g_j^0 \rightarrow g_j^{n+1} = \lambda g_j^n + (1-\lambda) g_j^n \rightarrow g_j \] 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 formulae used in carrying it out. (Online version in colour.)

functions are known and in which we controlled the level of noise and the distribution of data. This simple case also allows us to illustrate the efficiency of the procedure and the accuracy that can be obtained according to the quality of the data.

Figure 10. The colour map of a real function of two variables \( a(x, y) \) going from \([0, L] \times [-\pi, \pi] \) to \( \mathbb{R} \), for two levels of noise
\[ a(x, y) = \sqrt{2e^{-x^2/\sigma^2}} \sin y, \]
with \( \sigma = 0.1 \) and \( L = 2 \). The map is built as follows. We first discretize the 2D space \( (x, y) \) in \( \Omega = I \times J \) rectangular cells

\[ f_i g_j = a_{ij}, \quad i = 1, \ldots, I, \quad j = 1, \ldots, J. \] (3.11)

The values of \( f_i \) and \( g_j \) are unknown for all \( i \) and \( j \). 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 (3.7). Following the procedure to reduce this overdetermined system, we write the error function \( \Delta(X) \), for

**Figure 9.** Construction of the error function \( \Delta(X) \) in a simple two-dimensional case where \( a(x, y) = f(x) g(y) \). After discretizing 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 \( e_{ij} \) 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, \( (i, j) \), and summed up along all boxes. Starting from an initial guess of \( I + J \) values \( f_i, g_j \), \( i = 1, \ldots, I, \quad j = 1, \ldots, J \), 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 \( \Delta(X) \) is reached. The final values \( f_i, g_j \) (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 discretization index.
Box 1. Minimizing the error between data and a product form.

Let us consider an $N \times M$ two-dimensional array $a_{ij}$, for instance, built after averaging some quantity obtained from experiments on an $N \times M$ grid. We look for the best approximation of $a_{ij}$ under a product or separable form, $a_{ij} \approx x_i \times y_j$, $i = 1, \ldots, N$, $j = 1, \ldots, M$ (see the system (3.7), in the more general case when more than two dimensions are involved). Note that in the product form, the number of unknown variables is only $N + M$, compared with 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 $\Delta$:

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

(6.1)

Here, $e_{ij}$ 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 $x_1, \ldots, x_N, y_1, \ldots, y_M$ 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 $\nabla \Delta$, given by the partial derivatives of $\Delta$ with respect to its components; see the Finding minima in the boxed inset 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 $F: \mathbb{R}^D \to \mathbb{R}$ is the gradient vector $\nabla F(x)$, whose components are given by the partial derivatives of the function with respect to each component of $x$: $\nabla F(x) = (\partial F/\partial x_1, \partial F/\partial x_2, \ldots, \partial F/\partial x_D)$. The gradient vector points in the direction of maximum variation of $F$, and is therefore zero (i.e. equal to the null vector 0) when $x$ is an extremum, and in particular, a minimum:

$$\frac{\partial F}{\partial x_1} = 0, \quad \frac{\partial F}{\partial x_2} = 0, \ldots, \quad \frac{\partial F}{\partial x_D} = 0.$$  

(6.2)

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 generally satisfied for squared error functions like the one considered here.

Differentiating (6.1) with respect to $x_k$ gives

$$\frac{\partial \Delta}{\partial x_k} = \sum_{i=1}^{N} \sum_{j=1}^{M} \frac{\partial}{\partial x_k} \left[ e_{ij}(x_i y_j - a_{ij})^2 \right] = 2 \sum_{i=1}^{N} e_{ij} y_j \frac{\partial x_i}{\partial x_k} (x_i y_j - a_{ij}),$$  

(6.3)

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

$$\frac{\partial \Delta}{\partial x_k} = 2x_k \sum_{j=1}^{M} e_{ij} y_j^2 - 2 \sum_{j=1}^{M} e_{ij} y_j a_{ij},$$  

(6.4)

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

$$x_i = \frac{\sum_{j=1}^{M} e_{ij} y_j a_{ij}}{\sum_{j=1}^{M} e_{ij} y_j^2}, \quad i = 1, \ldots, N, \quad y_j = \frac{\sum_{i=1}^{N} e_{ij} x_i a_{ij}}{\sum_{i=1}^{N} e_{ij} x_i^2}, \quad j = 1, \ldots, M.$$  

(6.5)

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

$$X = (f_1, \ldots, f_n, g_1, \ldots, g_m),$$

$$\Delta(X) = \sum_{i=1}^{n} \sum_{j=1}^{m} e_{ij}(f_i g_j - a_{ij})^2,$$  

(3.12)

which has the same form as the one shown in equation (6.1) of box 1. Figure 9 shows how $\Delta(X)$ is calculated as the sum of all the squared local errors in each box, denoted by the vertical green line between the yellow and green 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

$$\frac{\partial \Delta}{\partial f_i} = 2 \sum_{j=1}^{m} e_{ij} g_j (f_i g_j - a_{ij})$$  

(3.13)

and

$$\frac{\partial \Delta}{\partial g_j} = 2 \sum_{i=1}^{n} e_{ij} f_i (f_i g_j - a_{ij}),$$  

(3.14)

and they are equal to zero when

$$f_i = \frac{1}{\sum_{j=1}^{m} e_{ij} g_j a_{ij}} / \sum_{j=1}^{m} e_{ij} g_j^2$$  

(3.15)

and

$$g_j = \frac{1}{\sum_{i=1}^{n} e_{ij} f_i a_{ij}} / \sum_{i=1}^{n} e_{ij} f_i^2.$$  

(3.16)
Note finally that the minus sign that appears in equation (3.9) can be dropped. For smooth and in a reasonable computational time. See also box 2.

chosen with the compromise that the iterations progress vice versa. Here, 

\[ x_{n+1} = \lambda x_n + (1 - \lambda) f(x_n), \quad (6.6) \]

where \( \lambda \in (0, 1) \) is the weight of the previous iteration in the value of the new iteration, averaged with what would have been the new iteration.

Thus, starting from an initial guess of the solution, \( X^0 = (f_1^0, f_2^0, \ldots, f_N^0, g_1^0, g_2^0, \ldots, g_M^0) \), it is possible to obtain the values of the \( f_\alpha \)s and the \( g_\beta \)s after one iteration, that is, \( X^1 = (f_1^1, f_2^1, \ldots, f_N^1, g_1^1, g_2^1, \ldots, g_M^1) \), repeating this process iteratively, \( X^i \rightarrow X^{i+1} \rightarrow X^{i+2} \rightarrow \cdots \) until convergence, i.e., \( X^{n+1} \approx X^n \).

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

\[ f_i^{n+1} = \lambda f_i^n + (1 - \lambda) f_i^0 \]

and

\[ g_j^{n+1} = \lambda g_j^n + (1 - \lambda) g_j^0. \]

Here, \( f_i^n \) and \( g_j^n \) are calculated from the values of \( X^n \) in the previous step with the formulæ (3.15) and (3.16), and \( \lambda \in [0, 1] \) is chosen with the compromise that the iterations progress smoothly but in a reasonable computational time. See also box 2.

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

The final state to which the iterative method has converged, that is, the points \( (f_i^I)_{i=1}^I \) and \( (g_j^I)_{j=1}^J \) that solve the reduced system (6.5), is shown in figure 10c,d. There is an excellent agreement with the original data, despite the addition of noise, including in the case of a much larger noise than the one inferred in the actual data of our fish experiments.

Note that if \( f \) and \( g \) are a solution of (6.5), then the functions \( (1/\alpha)f \) and \( \alpha g \), where \( \alpha \) is a real number, are also a solution of (6.5), since the product of the two functions remains invariant. Hence, we need to impose an additional condition in order to account for this under-determination and to generally allow for a proper comparison of reconstructed interaction functions. Following [12], we chose to normalize all angular functions so that their squared average is equal to 1. In particular, \( g \) is normalized such that \( 1/(2m) \int g(y)^2 \, dy = 1 \), a normalization applied in figure 10c,d.

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

(c) On the hypothesis of separation of variables of interaction functions

The central assumption of the extraction procedure is the separation of variables made in (3.4)–(3.6). Without this hypothesis, the solution of equation (3.3) far from the wall is \( \delta h \), and with fewer nodes in the reconstruction procedure.

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

\[
\hat{a}(d, \psi) = \sin \psi \sum_{k=0}^{
fty} c_k(d) \cos (k \psi). \tag{3.19}
\]

In order to be separable as \( a(d, \psi) = f(d)g(\psi) \), the functions \( c_k(d) \) in (3.19) have all to be proportional to some \( f(d) \), that is, \( c_k(d) = a_k f(d) \), with \( a_k \) constant for all \( k \). Then, \( g(\psi) = \sin \psi \sum_{k=0}^{
fty} a_k \cos (k \psi) \).

A simple and biologically meaningful example, inspired by the fish interaction functions, would consist in writing \( c_k(d) = a_k \exp \left[ -d^2 / (2l_0^2) \right] \) in (3.19) and using only the modes \( k = 0 \) and 2:

\[
a_0 = \sin \psi \left[ a_0 e^{-d^2 / (2l_0^2)} + a_2 e^{-d^2 / (2l_0^2)} \sin (2 \psi) \right].
\]

When \( l_0 = l_0 \), the variables are separable and we have

\[
a_0 = a_0 \sin \psi \left[ 1 + a_2 \cos (2 \psi) \right], \tag{3.20}
\]

which, when \( a_2 = \sqrt{2} l_0 = r_0 / \sqrt{2} \) and \( a_2 = 0 \), corresponds to the simple case used in \( \S \) 3b, for which the reconstruction procedure worked quite well.

Let us 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 (3.19), we consider two examples \( b_1(d, \psi) \) and \( b_2(d, \psi) \) of the form

\[
b_{1,2}(d, \psi) = \sqrt{2} e^{-d^2 / (2l_0^2)} \sin \psi \left[ + 1.13 e^{-d^2 / (2l_0^2)} \sin (2 \psi) \right],
\]

where \( l_0 = 2 l_0 / 3 \approx 0.047 \) and \( l_0 = 3 l_0 / 2 \approx 0.106 \), respectively.

Figure 10. Force maps and reconstruction of the interaction functions in a simple case. (a, b) Force maps of the function \( f(x, y) + \eta(x, y) \) 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 \( x \) as the one observed in the experiments with \( H. rhodostomus \). (c, d) Reconstruction of the interaction functions \( f(x) \) and \( g(y) \) extracted with our procedure (red dots and blue circles), compared with the analytical expressions \( \exp(-\alpha x^2) \) 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 fewer nodes in \( x \), 11 instead of 20. Note that the function \( g(y) \) is normalized.

\( \delta \) hypothesis, the solution of equation (3.3) far from the wall is always expandable in a Fourier series as...
and $\alpha_2 = 0.8a_0 \approx 1.13$ in both functions. Note that the first term is precisely the function used in the simple case of §3b, and the presence of the second term makes the functions $b_1$ and $b_2$ not separable.

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

$$f(d) = \beta_8 e^{-d/[2l^2k^2]}$$

and

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

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

$$f(d)g(\psi) = 1.37e^{-d/[2l^2k^2]} \sin \psi [1 + 0.62 \cos(2\psi)],$$

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

Although the procedure provided an apparently satisfactory result (figure 12a,b are quite similar), the reconstructed functions fail to reproduce some features such as the (two) changes of variation of the intensity along the radial coordinate (first increasing, then decreasing) when the neighbour is at one side of the focal individual ($\varphi \approx \pm \pi/2$), and introduce an artificial slight decrease in the intensity when the neighbour is far away ($d \approx 3$ BL) and at one side ($\varphi \approx \pm \pi/2$) of the focal individual (figure 12c,d).

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

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

In both examples, the worst agreement between the reconstructed interaction and the actual one is obtained for the angular dependence of the interaction at very large distance (blue curves in figures 12d and 13d), and for the radial dependence of the interaction at $\psi = \pm \pi/2$ (red curves in figures 12c and 13c). However, both cases correspond to situations where the interaction is the weakest and the reconstructed interaction matches perfectly the actual one in situations where the interaction is most significant (black and red curves in both panels (d); black and blue curves in both panels (c)).

Let us now discuss the general validity of the product assumption on the illustrative example of the interaction of a fish with a circular wall. In this case, the heading angle change between two time steps (kicks) is $\delta\varphi(r_\psi \theta_\psi)$, where $r_\psi$ is the distance of the fish from the wall, and $\theta_\psi$ is the angle between the fish heading and the normal to the wall. Here, the product hypothesis assumes to assuming that $\delta\varphi(r_\psi \theta_\psi) = f(r_\psi)g(\theta_\psi)$. As explained in [12], the product assumption is verified for physical particles interacting with a wall via a conservative force (deriving from a potential energy), with $g(\theta_\psi)$ given exactly by $\sin \theta_\psi$ (projection of the central force on the normal to the velocity, i.e. on the angular acceleration). For animals, and in particular fish, the anisotropic perception of their environment generally leads to non-conservative interactions. The product hypothesis assumes that this anisotropic perception can be fully encoded in a non-trivial function $g(\theta_\psi)$ generally different from a simple sinus, while the function remains odd if the left/right symmetry is preserved. As explained above, $\delta\varphi(r_\psi \theta_\psi)$ for animals takes the general Fourier expansion form of equation (3.19), $\delta\varphi(r_\psi \theta_\psi) = \sin(\theta_\psi) \sum_{c_{2k-1}} c_k(r_\psi) \cos(\theta_\psi)$. Strictly speaking, $\delta\varphi(r_\psi \theta_\psi)$ can be separated into the product of two functions of $r_\psi$ and $\theta_\psi$ only if all functions $c_k(r_\psi)$ are proportional to each other. However, if these functions decay similarly with the distance to the wall $r_\psi$, the proportionality assumption and hence the product assumption would be only weakly invalid, as illustrated in the practical examples presented above. In fact, we generally expect only a few Fourier modes to be relevant in the above expansion (in [12], it was found that only one or two 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 $c_k(r_\psi)$ 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 $r_\psi$. 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

![Figure 11](https://example.com/figure11.png)

**Figure 11.** Evolution of the iterative process for the minimization of the error function $\Delta f$ in the simple case where $a(x, y) = \sqrt{2 \exp[-(x^2+y^2)]} \sin y$. Blue circles denote the successive values of the $l+1$ components of the unknown functions $f^0$ and $g^0$ during the first 120 steps of the iterative process. Red lines show the initial guess of each function $f^0_j = g^0_j = 1, i = 1, \ldots, l$ and $j = 1, \ldots, l$, and the final state of convergence. Note that $n$ is the index of iteration, not the time discretization index. (Online version in colour.)
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 §3 is based on the assumption that social interactions are combined in an additive form (see step 3, figure 6). In equation (3.2), two functions \( \delta \varphi_{\text{All}} \) and \( \delta \varphi_{\text{All}} \) 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 equations (3.5) and (3.6). Steps 4.1–4.3 in figure 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 figure 6) is used to say that \( \delta \varphi \) must change sign when \( \psi_{ij} \)
and $\phi_i$ both change sign, and that, consequently, the same must happen for $\delta\phi_{\text{All}}$ 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 $\delta\phi_{\text{All}}$ must be an odd function of $\psi_{ij}$, and thus an even function of $\phi_i$, 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 $\phi_i$, and thus an even function of $\psi_{ij}$.

This allows us to rewrite the interaction functions from equations (3.5) and (3.6) as follows:

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

and

$$\delta\phi_{\text{All}}(d_{ij}, \psi_{ij}, \phi_i) = F_{\text{All}}(d_{ij}) E_{\text{All}}(\phi_i) O_{\text{All}}(\psi_{ij})$$

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 §3).

Figure 14 shows the social interaction functions reconstructed with the procedure described in the previous section in the case of two $H. rhodostomus$ (figure 14a–c) and two $D. rerio$ (figure 14d–f) 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

![Figure 13.](image-url)
are very different. Hence, we used the analytical expressions introduced in [12] for *H. rhodostomus* (solid lines in figure 14) to fit the discrete values of *D. rerio* extracted from the experimental data in step 4 of our method (figure 6), that is, with the procedure described in §3 (points in figure 14).

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

\[ F_{\text{Att}}(d) = \gamma_{\text{Att}} \left( \frac{d}{d_{\text{Att}}} - 1 \right) \frac{1}{1 + \left( \frac{d}{d_{\text{Att}}} \right)^2} \]  \tag{4.3}

and

\[ F_{\text{Ali}}(d) = \gamma_{\text{Ali}} \frac{d}{d_{\text{Ali}}} \exp \left[ - \left( \frac{d}{l_{\text{Ali}}} \right)^2 \right], \]  \tag{4.4}

where the values of the parameters depend on the species and on the size of the arena. Note that the expression for 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, \( l_{\text{Ali}} \) and \( l_{\text{Ali}} \) are the ranges of each interaction (the higher the value, the longer the range), and \( d_{\text{Att}} = 1 \) BL is a characteristic length used to make \( \gamma_{\text{Att}} \) dimensionless. Having dimensionless factors \( \gamma_{\text{Att}} \) and \( \gamma_{\text{Ali}} \) allows direct comparison between intensities of social interactions and the intensity of the interaction with obstacles \( \delta_{\text{wall}} \) and the spontaneous decision term \( \delta_{\phi_{\text{w}}} \) in equation (3.1). Table 1 shows the parameter values corresponding to each species. Note that the values for *H. rhodostomus* have been adapted from those in [12], according to the slightly different expression used here.

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

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

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

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

Following step 5 of our methodology (figure 6), these analytical expressions should be implemented in the model introduced in §3. Then, numerical simulations of the model should be performed and compared with the known experiments, and predictions should be made, which must be verified *a posteriori*. This corresponds to step 6 of our methodology (figure 6); this has been done for *H. rhodostomus* in [12], and will be done elsewhere for *D. rerio* and other species.

Comparing the interaction functions found for both species, we observe that they have a similar shape, especially
the angular functions (see the angular functions of attraction in figure 14b, for which one can use the same function of the angle of perception $\psi_i$). The intensities of attraction and alignment have the same order of magnitude in both species: the maximum of the attraction is around 0.4 in both species, and the maximum of the alignment is around 0.2–0.3 (figure 14a).

The most important difference between the two species is that the range of the interactions is much larger in *H. rhodostomus* than in *D. rerio*: in *H. rhodostomus*, the maximum intensities of attraction and alignment are around 7 and 3.5 BL respectively, while in *D. rerio* these maxima are both near 1.5 BL. Moreover, the intensity of these interactions decays more rapidly in *D. rerio* than in *H. rhodostomus*, especially with respect to the fish BL (in *D. rerio*, the alignment intensity is zero beyond 5 BL (approx. 22.5 cm), but it is still noticeable at a distance of 10 BL (approx. 30 cm) in *H. rhodostomus*). Attraction almost always dominates alignment in *D. rerio* (the intersection of the red and blue lines in figure 14a is at around 0.5 BL), while, in *H. rhodostomus*, alignment was found to dominate at short distances (under 2.5 BL).

In *H. rhodostomus*, a fish $i$ is subject to a stronger attraction when the other fish $j$ is at its right or left side ($O_{Ai}$ reaches its highest values when $\psi_{ij} \approx \pm 90^\circ$; see figure 14b) and moves more or less perpendicular to it ($E_{Ai}$ is higher when $\phi_{ij} \approx 90–100^\circ$; see figure 14c). Alignment is stronger when the other fish is in front ($E_{Ai}$ is higher when $|\psi_{ij}| < 80^\circ$). In *D. rerio*, attraction is stronger when the other fish is clearly behind the focal fish ($O_{Ai}$ is higher when $\psi_{ij} \approx 135^\circ$) and moves in the opposite direction ($E_{Ai}$ is higher when $|\phi_{ij}| > 100^\circ$; see figure 14c). In *D. rerio*, the strength of the alignment vanishes when fish are already almost aligned ($O_{Ai} = 0$ when $|\psi_{ij}| < 30^\circ$; see figure 14c). In *D. rerio*, the strength of the alignment is active essentially when both fish are perpendicular to each other (the high intensity of $|O_{Ai}| \approx 1.5$ is reached when $\phi_{ij} \approx \pm 85^\circ$) and the focal fish has its neighbour at one of its sides ($E_{Ai}$ peaks at $\psi_{ij} \approx \pm 85^\circ$), while in *H. rhodostomus*, the ranges of interaction both in the angle of relative heading and in the angle of perception of the neighbour are much wider: alignment is active when the neighbour is ahead of the focal fish ($1.5 < |\psi_{ij}| < 100^\circ$) and fish are simply slightly aligned ($45^\circ < |\phi_{ij}| < 135^\circ$).

In summary, when swimming in pairs, *H. rhodostomus* interact in a much wider range of situations than *D. rerio*. This is true with respect to the three state variables of a focal fish: (1) the distance $d_{ij}$ at which both attraction and alignment interactions are active is much larger in *H. rhodostomus* than in *D. rerio*; (2) the zone around a focal fish where the strength of the interactions with a neighbour is important is much wider in *H. rhodostomus* than in *D. rerio*, and (3) the same is true for the range of relative headings for which the intensity of interactions between fish is not negligible. Finally, the maximum intensity of the interaction is similar in both species, although high-intensity values are reached in a much wider range of situations in *H. rhodostomus* than in *D. rerio*.

5. Discussion and conclusion

Behavioural biology has recently become a ‘big-data science’ mainly supported by the advances in imaging and tracking techniques. These new tools have revolutionized the observation and quantification of individual and collective animal behaviour, improving to unprecedented levels the variety and precision of available data [9,26–28]. 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 qualitatively different from those of their individual members, and whose behaviours are impossible to predict from *a priori* 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 investigation of 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 that 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 etc.) and another individual (distance between individuals, relative velocities, viewing angle etc.). In addition, each interaction component (repulsion, attraction, alignment etc.) 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 figure 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 parameter, 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 (figure 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 the predictive power of the model to be assessed. 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 five to eight 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 five 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 exploits 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 general qualitative conclusions to be made 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 that can be exploited to understand and explain diverse experimental features and various forms of collective behaviour, and that 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 are solely and fully encoded in their different measured interaction functions.

In the specific cases 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 (figure 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 $|\psi_i|<30^\circ$). Both features contribute to a much weaker coordination of the motion in groups of two fish in *D. rerio*, compared with *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 (figure 14d), 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 $\phi_0=0$ in *D. rerio*.
(figure 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*.

6. Material and methods

(a) 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 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.

(b) Study species

*Hemigrammus rhodostomus* were purchased from Amazonie Labège (http://www.amazonie.com) in Toulouse, France. Fish were kept in 1501 aquariums on a 12:12 h, dark:light photoperiod, at 26.8°C (±1.6°C) and were fed ad libitum with fish flakes. The average BL of the fish used in the experiments was 31 mm. Sixty wild-type *D. rerio* with short fins (AB strain) were acquired from a pet shop, and stored in a 60 l aquarium. The average BL of the fish used in the experiments was approximately 4.5 cm. The water in the housing aquarium was kept at a temperature of 26°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.

(c) Experimental procedures and data collection

The experimental tank (120 × 120 cm) 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 set-up, placed in a chamber made by four opaque white curtains, was surrounded by four LED light panels giving isotropic lighting. A circular tank of radius *R* = 25 cm was set inside the experimental tank filled with 7 cm of water of controlled quality (50% water purified by reverse osmosis and 50% water treated by activated carbon) heated to 26.1°C (±0.3°C). Reflections of light due to the bottom of the experimental tank were avoided thanks to a white PVC layer. Each trial started by placing two fish randomly sampled from their breeding tank in a circular tank. Fish were left for 10 min to habituate before the start of the trial. A trial consisted in 1 or 3 h 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 × 1080 pixels).

In the case of *D. rerio*, the experimental set-up had dimensions of 100 × 100 × 25 cm, inside which a circular tank of radius *R* = 29 cm was placed. The sides and bottom part of the tank were covered with a Teflon plate to avoid reflections. Furthermore, the set-up was confined behind white sheets to isolate the fish from external stimuli in the room, while also maintaining a consistent lighting environment inside the set-up bounds. A uniform luminosity for the room was provided by four 110 W fluorescent lamps placed at each of the four sides of the tank. Prior to placing fish in the experimental set-up, 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 set-up. 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 reinserted 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 metres 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 moved less than 0.5 BL per second during 4 s were removed. More than 8 h remained during which one fish could kick.

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

(d) 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*<sub>sw</sub>, *t* + *t*<sub>sw</sub>, where *t*<sub>sw</sub> is a time window of 0.32 s. 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.

(e) Fortran code

We provide the code to perform the analysis described in §3b, together with a simple script for quick visualization of the result in the gnuplot environment and the required data in Figshare: https://doi.org/10.6084/m9.figshare.11777325.

Data accessibility. Supporting data are available from the Figshare Repository: https://doi.org/10.6084/m9.figshare.9933356. We provide the code to perform the analysis described in §3.2, together with a simple script for quick visualization of the result in the gnuplot environment and the required data in Figshare: https://doi.org/10.6084/m9.figshare.11777325.

Authors’ contributions. R.E., V.L., V.P. and F.B. performed research; R.E., V.L., C.S. and G.T. analysed data. R.E., V.L., C.S. and G.T. wrote the paper with an input from all other authors.

Competing interests. We declare we have no competing interest.

Funding. 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.

Acknowledgements. 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.

Endnotes

1. The average function of *a*(x, y) with respect to the variable *x* is given by *(a)*<sub>y</sub>*y* = ∫ *p*(x, *y*) *d*<sub>x</sub>, where *p*(x, y) is the probability of occurrence of the state (*x*, y). However, force maps calculate the *(a)*<sub>y</sub>*y* as the mean of the values of *a*(x, y) over all the values of *x*, i.e. *(a)*<sub>y</sub> = ∑ *a*(x, y) / ∑ 1, as if *p*(x, y) = 1 for all (x, y), because knowing *p*(x, y) is part of the problem.

2. The function *f* must be contractive and the initial value of the iterations must be sufficiently close to the fixed point.