A statistical method for identifying different rules of interaction between individuals in moving animal groups – Supporting Information

T M Schaerf[1,*] , J E Herbert-Read[2, 3,*], and A J W Ward[4]

[1]School of Science and Technology, University of New England, Armidale, NSW 2351, Australia

[2]Department of Zoology, University of Cambridge, Cambridge CB2 3EJ, UK

[3]Aquatic Ecology, University of Lund, Lund, 223 62, Sweden

[4]Animal Behaviour Lab, School of Life and Environmental Sciences, University of Sydney, NSW 2006, Australia

[*]Corresponding author contact details:

Timothy Schaerf

Booth Block C027

School of Science and Technology

University of New England, Armidale, NSW 2351, Australia

Email: tschaerf@une.edu.au

James Herbert-Read

Department of Zoology

University of Cambridge, Downing St, Cambridge, CB2 3EJ, United Kingdom

Email: jh2223@cam.ac.uk

S1 Estimation of rules of interaction and related quantities

S1.1 Preliminary calculations and fundamental quantities

Figure S1 illustrates some of the key measures of locomotion, relative neighbour positions, and relative neighbour alignment that are derived and used throughout this section and the next. The ultimate goal of the calculations is to estimate how individuals adjust their velocity in response to (or as a function of) the relative positions of their neighbours.

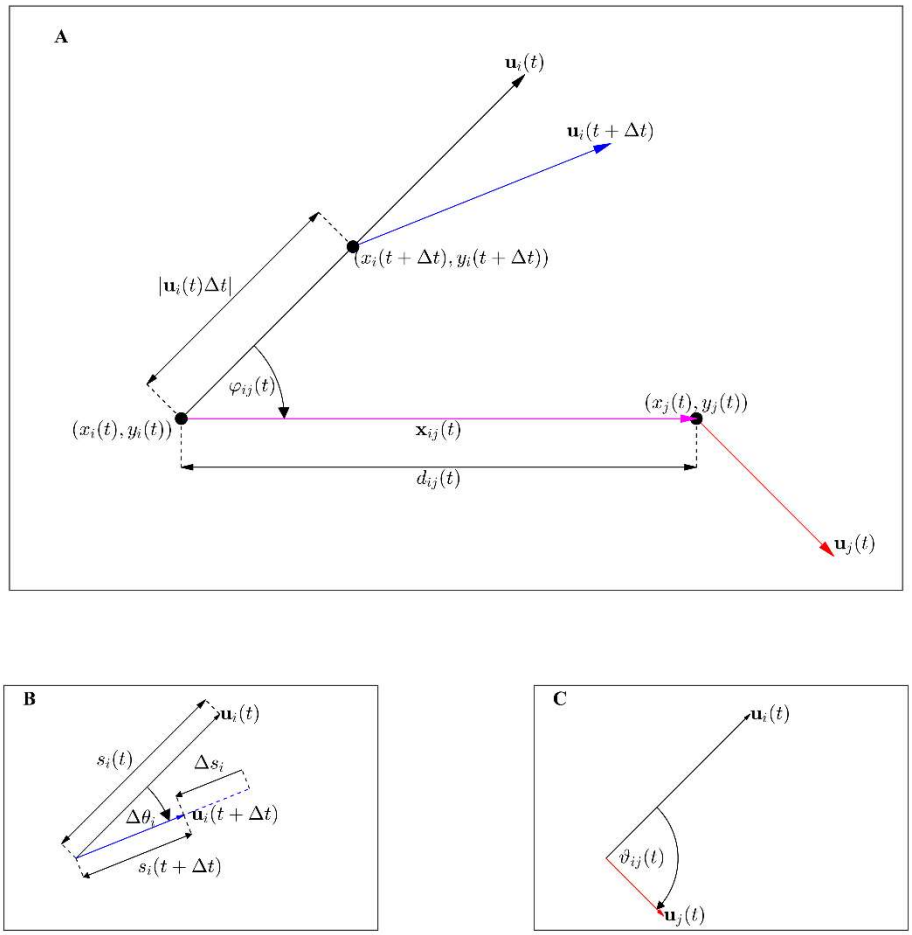

Figure S1: Quantities associated with locomotion, relative neighbour locations and interactions. A: $(x_i(t), y_i(t))$ (at the location of the nearby black dot) and $\mathbf{u}_i(t)$ (black single-headed arrow) denote the position and velocity of a *focal* individual *i* at some discrete time *t*. $(x_i(t+\Delta t), y_i(t+\Delta t))$ and $\mathbf{u}_i(t+\Delta t)$ (blue arrow) represent the same quantities at the next discrete time, $t+\Delta t$, where Δt is the constant time between tracking outputs. It is assumed that individual *i* moves directly along the straight line segment joining $(x_i(t), y_i(t))$ and $(x_i(t+\Delta t), y_i(t+\Delta t))$ between times t and $t+\Delta t$, which is a distance of $|\mathbf{u}_i(t)\Delta t|$ units. A neighbouring individual, indexed *j*, has coordinates $(x_j(t), y_j(t))$ and velocity $\mathbf{u}_j(t)$ (red arrow) at time t. The vector pointing from the position of individual *i* at time *t* to the position of individual *j* at the same time is denoted $\mathbf{x}_{ij}(t)$ and has length $d_{ij}(t)$ (this is the linear distance between individuals *i* and *j*). The internal angle between $\mathbf{u}_i(t)$ (which indicates the direction of motion of individual *i*) and $\mathbf{x}_{ij}(t)$ (along the straight line segment between individuals *i* and *j*) is denoted $\varphi_{ij}(t)$, and is measured using the convention that clockwise rotations from $\mathbf{u}_i(t)$ to $\mathbf{x}_{ij}(t)$ correspond to negative angles, and anticlockwise rotations

correspond to positive angles (this is a standard mathematical convention). B: This panel illustrates quantities associated with the change in velocity of individual *i*. The instantaneous speed of the individual at time *t* is denoted $s_i(t)$, and is the magnitude (or length) of the individual's velocity vector, $\mathbf{u}_i(t)$ (black arrow). Similarly, the speed of the individual at time $t + \Delta t$ is $s_i(t + \Delta t)$, so that the difference in speed of the individual from one time step to the next is $\Delta s_i = s_i(t + \Delta t) - s_i(t)$ (a quantity that will be positive if the individual speeds up, or negative if the individual slows down). In this example, Δs_i is negative. The change in the direction of motion of the individual is measured via the internal angle $\Delta\theta_i$, which is formed when the vectors $\mathbf{u}_i(t)$ and $\mathbf{u}_i(t + \Delta t)$ are placed tail-to-tail, as illustrated. Similar to A, the convention applied is that clockwise rotations from $\mathbf{u}_i(t)$ to $\mathbf{u}_i(t + \Delta t)$ are identified as negative, whereas anticlockwise rotations are positive. C: The direction of motion of individual *j* (red arrow) relative to that of individual *i* (black arrow) at time *t* is measured via the internal angle $\vartheta_{ij}(t)$, formed when $\mathbf{u}_i(t)$ and $\mathbf{u}_j(t)$ are placed tail-to-tail. Again the anticlockwise is positive convention is applied for identifying the sense of the associated rotation. Further details on the derivation and use of the quantities illustrated in the above panels are given in sections S1.1 and S1.2.

Given time series of the coordinates of individuals in two dimensions, obtained by some tracking method, and usually as part of multiple separate sets of observations, we first calculate fundamental measures of individual locomotion as follows. We smooth the x and y components of each individual's track using a Savitzky-Golay filter (implemented through MATLAB's intrinsic *smooth* function with span 5 and degree 2; a number of other smoothing methods would be valid to apply in place of this filter) prior to all diagnostic calculations, to take into account the potential presence of small noisy variations in position in the data that are a result of small inaccuracies in the tracking. With a span of 5, the Savitzky-Golay filter obtains each smoothed data point based on 5 raw data points which includes the two data points immediately preceding the current data point, the current data point, and the two data points immediately following the current data point. The data for the case study later in this paper was recorded at 40 frames per second, and thus smoothed points for a particular discrete time were obtained using data from 0.05 seconds before the current time to 0.05 seconds after the current time.

In the following descriptions, vectors are typeset in bold font, **i**, **j**, and **k** represent unit vectors parallel to the positive *x*-, *y*-, and *z*-axes respectively, the $\bullet$ symbol denotes the scalar dot/inner product of two vectors, and the $\times$ denotes the vector cross product of two vectors. Writing $\left(x_i(t), y_i(t)\right)$ as the coordinates of individual i at time t after smoothing, we determine the x and y components of that individual's velocity using the standard forward-difference approximations:

$$u_i(t) = \frac{x_i(t + \Delta t) - x_i(t)}{\Delta t}, \quad v_i(t) = \frac{y_i(t + \Delta t) - y_i(t)}{\Delta t}, \tag{1}$$

respectively, where Δt is the constant duration between consecutive video frames or GPS outputs. An individual's speed at time t is then approximated as:

$$s_i(t) = \sqrt{\left(u_i(t)\right)^2 + \left(v_i(t)\right)^2}. \tag{2}$$

Following immediately from this calculation we determine the change in an individual's speed over time via:

$$\frac{\Delta s_i}{\Delta t}(t) = \frac{s_i(t+\Delta t) - s_i(t)}{\Delta t}. \tag{3}$$

(The above measure is referred to as tangential acceleration in (1).) The measure in equation (3) differs from both the acceleration of an individual (a vector), and the magnitude of acceleration. $\Delta s_i / \Delta t$ can take negative values (representing slowing), so it is more illuminating to examine than the magnitude of acceleration (which is non-negative by definition) when it is of interest to determine when individuals are speeding up or slowing down.

To determine the change in direction over time of an individual we perform the following sequence of calculations. We first construct unit vectors, $\hat{\mathbf{u}}_i(t) = \hat{u}_i(t)\mathbf{i} + \hat{v}_i(t)\mathbf{j}$, in the direction of each individual's velocity vector, with components

$$\hat{u}_i(t) = \frac{u_i(t)}{s_i(t)}, \quad \hat{v}_i(t) = \frac{v_i(t)}{s_i(t)}. \tag{4}$$

The internal angle between an individual's velocity vectors at times t and $t + \Delta t$ can then be determined via the dot product:

$$\alpha_i(t) = \cos^{-1}\left(\hat{\mathbf{u}}_i(t) \bullet \hat{\mathbf{u}}_i(t+\Delta t)\right). \tag{5}$$

The sense of rotation associated with the above change in direction in the xy-plane can be deduced with the help of a cross product via

$$\beta_i(t) = \operatorname{sgn}\left(\left(\hat{\mathbf{u}}_i(t) \times \hat{\mathbf{u}}_i(t+\Delta t)\right) \bullet \mathbf{k}\right), \tag{6}$$

where $\operatorname{sgn}(A)$ is 1 if A is positive, -1 if A is negative, and 0 if A is zero. For anticlockwise rotations $\beta_i(t)$ is positive, whereas for clockwise rotations $\beta_i(t)$ is negative. Using equations (5) and (6), one way to then estimate the change in direction of motion over time is:

$$\frac{\Delta \theta_i}{\Delta t}(t) = \begin{cases} \dfrac{\beta_i(t)\alpha_i(t)}{\Delta t} & \text{if} \quad \beta_i(t) \neq 0, \\ \dfrac{\alpha_i(t)}{\Delta t} & \text{if} \quad \beta_i(t) = 0, \end{cases} \tag{7}$$

such that anticlockwise turns/turns to the left relative to an individual's direction of motion are associated with positive $\Delta\theta/\Delta t$, and clockwise turns/turns to the right are associated with negative $\Delta\theta/\Delta t$.

S1.2 Determination of average rules of interaction

The approach used by (1-3) to extract rules of interaction from trajectory data, and the approach described here, is to use the data to construct functions that describe how an

individual adjusts the components of its velocity on average as a function of the relative coordinates of groupmates, in a frame of reference where the individual has a consistent direction of motion. The complementary methods developed by (2) and (1) used slightly different, but nevertheless consistent, formulations to examine changes in velocity. Katz et al. (2011) described changes in velocity via the components of acceleration parallel and perpendicular to the direction of motion of an individual (termed the speeding force and turning force), and constructed plots of their fitted functions such that the individual adjusting its motion was located at the origin, moving parallel to the positive vertical axis. For this study we used the alternative approach developed by (1) and refined in (3), where changes in velocity are described via changes in speed over time and changes in direction of motion over time (as described by equations (3) and (7) above), with the individual located at the origin and travelling parallel to the positive x-axis in the plots generated as a result of the analysis which follows.

Having calculated all the quantities in the section S1.1, the next step is to calculate the coordinates of all individuals j relative the coordinates of individual i (which will be referred to as the focal individual in parts of the following text). The way that we do this, is by first determining the linear distance from individual i to its partner j via the standard distance formula:

$$d_{ij}(t) = \sqrt{\left(x_j(t) - x_i(t)\right)^2 + \left(y_j(t) - y_i(t)\right)^2}. \tag{8}$$

We then calculate the angle between the direction of motion of the focal individual, i, as given in component form by equations (4), and the directed straight line segment from the coordinates of i to the coordinates of their groupmate j. The unit vector from the position of individual i to the position of individual j has components

$$\hat{x}_{ij}(t) = \frac{x_j(t) - x_i(t)}{d_{ij}(t)}, \quad \hat{y}_{ij}(t) = \frac{y_j(t) - y_i(t)}{d_{ij}(t)}, \tag{9}$$

and thus can be written in vector form as $\hat{\mathbf{x}}_{ij}(t) = \hat{x}_{ij}(t)\mathbf{i} + \hat{y}_{ij}(t)\mathbf{j}$.

Following an analogous set of calculations to those used to calculate changes in direction in equations (5), (6), and (7), the internal angle between the direction of motion of individual i, and the straight line segment from individual i to groupmate j, denoted by $\varphi_{ij}(t)$, is then calculated via

$$\varphi_{ij}(t) = \begin{cases} \lambda_{ij}(t)\phi_{ij}(t) & \text{if} \quad \lambda_{ij}(t) \neq 0, \\ \phi_{ij}(t) & \text{if} \quad \lambda_{ij}(t) = 0, \end{cases} \tag{10}$$

where the magnitude of the associated angle is given by

$$\phi_{ij}(t) = \cos^{-1}\left(\hat{\mathbf{u}}_i(t) \bullet \hat{\mathbf{x}}_{ij}(t)\right), \tag{11}$$

and the sense of rotation (anticlockwise or clockwise) associated with the angle is given by

$$\lambda_{ij} = \operatorname{sgn}\left(\left(\hat{\mathbf{u}}_i(t) \times \hat{\mathbf{x}}_{ij}(t)\right) \bullet \mathbf{k}\right), \tag{12}$$

where $\lambda_{ij}(t) = 1$ coincides with an anticlockwise rotation, and $\lambda_{ij}(t) = -1$ corresponds to a clockwise rotation ($\lambda_{ij} = 0$ if $\hat{\mathbf{u}}_i(t)$ and $\hat{\mathbf{x}}_{ij}(t)$ are parallel). Combined, $d_{ij}(t)$ and $\varphi_{ij}(t)$ describe the position of individual *j* relative to that of individual *i* in a polar coordinate system where individual *i* is located at the origin, with velocity parallel to the positive x-axis. In Cartesian coordinates, the coordinates of individual *j* relative to individual *i*'s position and direction of motion are:

$$\begin{aligned} x_{ij,\,\text{relative}}(t) &= d_{ij}(t)\cos\left(\varphi_{ij}(t)\right), \\ y_{ij,\,\text{relative}}(t) &= d_{ij}(t)\sin\left(\varphi_{ij}(t)\right). \end{aligned}$$

(Note that since $\left(x_{ij,\,\text{relative}}(t), y_{ij,\,\text{relative}}(t)\right)$ gives the coordinates of individual *j* relative to both the location and direction of motion of individual *i*, in general there will not be any symmetry between $\left(x_{ij,\,\text{relative}}(t), y_{ij,\,\text{relative}}(t)\right)$ and $\left(x_{ji,\,\text{relative}}(t), y_{ji,\,\text{relative}}(t)\right)$.)

We then divide the domain centred on the focal individual, *i*, into a set of overlapping square bins. To estimate the mean change in speed over time of an individual as a function of the relative (x, y) coordinates of groupmates, we then deposit $\frac{\Delta s_i}{\Delta t}(t)$ into all the bins that contain the point $\left(x_{ij,\,\text{relative}}(t), y_{ij,\,\text{relative}}(t)\right)$, repeating this for all discrete time steps t, all groupmates j, treating all individuals in turn as the focal individual, and aggregating data across all sets of observations. For the case study of mosquitofish pairs, the overlapping bins were constructed such that the bins covered the domain $-100 < x \le 100$, $-100 < y \le 100$ (in millimetres), with square bins with side length 16 mm (a little over half a body length), and with the left and bottom edges of consecutive bins separated by 4 mm.

After the data is binned, we determine the mean value within each bin, with the result being an estimate of the average change in speed of an individual as a function of the relative coordinates of its groupmates. We then visualise the resulting function with an appropriate MATLAB function, such as the *surf* command. A similar process is used to determine the average change in direction of motion of an individual over time, with the only difference being that $\frac{\Delta \theta_i}{\Delta t}(t)$ values are binned instead of $\frac{\Delta s_i}{\Delta t}(t)$.

The process above can be modified to examine how other measures of locomotion correlate with the relative coordinates of groupmates by binning the measure of interest (for example the mean speed of individuals as a function of the relative coordinates of groupmates (as in (3, 4)) or differences in speed between individuals and groupmates as a function of the relative coordinates of groupmates (4)). The method of binning can also be modified to take into account more, fewer, or different independent variables, for example to examine changes in velocity as a function of both the relative (x, y) coordinates of groupmates and the speed of focal individuals (1-3) or the speed of groupmates (2), or the components of changes in velocity as a function of only one of the relative *x*- or *y*-coordinates of groupmates (3).

The process can also be modified to give details about local group structure through the probability of groupmates occupying particular relative (x, y) coordinates, denoted $p(x, y)$, and measures of the relative alignment between focal individuals and their groupmates (such calculations were immediate predecessors to the rules of interaction calculations described in

(1, 2), as applied to data from observations of surf scoters in (5)). Estimates of the local probability density function for the positioning of groupmates, p, are the most straightforward to construct, requiring tallies of how often groupmates occupy particular (x, y) coordinates relative to the focal individual, with these tallies then divided by the total number of points in all bins to determine a relative frequency. There are multiple methods for examining the relative alignment of individuals (see for example (3, 5)); the approach that we used for this study is as follows. Applying calculations very similar to the determination of the change in direction of motion described by equations (5), (6), and (7), and the relative direction from individual i to individual j as described by equations (10), (11), and (12), the direction of motion of individual j relative to that of individual i at time t is given by

$$\vartheta_{ij}(t) = \begin{cases} \rho_{ij}(t)\psi_{ij}(t) & if \quad \rho_{ij}(t) \neq 0, \\ \psi_{ij}(t) & if \quad \rho_{ij}(t) = 0, \end{cases} \tag{13}$$

where the magnitude in the angular differences between the directions of motion of the individuals is given by

$$\psi_{ij}(t) = \cos^{-1}\left(\hat{\mathbf{u}}_i(t) \bullet \hat{\mathbf{u}}_j(t)\right), \tag{14}$$

and the sense of rotation from $\hat{\mathbf{u}}_i(t)$ to $\hat{\mathbf{u}}_j(t)$ is given by

$$\rho_{ij}(t) = \mathrm{sgn}\left(\left(\hat{\mathbf{u}}_i(t) \times \hat{\mathbf{u}}_j(t)\right) \bullet \mathbf{k}\right). \tag{15}$$

As with other calculations described in this section, we then deposit all values of $\vartheta_{ij}(t)$ into all bins containing $\left(x_{ij,\,\mathrm{relative}}(t),\, y_{ij,\,\mathrm{relative}}(t)\right)$ for all pairings of focal individuals *i* and their groupmates *j*, and for all discrete times *t*. We then determine the means of the sets of angles contained in each bin, and the focus of each set of angles about the mean using standard methods of circular statistics (6). For a set of *M* angles, denoted ϑ_k, contained in a particular bin, the mean angle is given by $\overline{\vartheta} = \mathrm{atan2}(Y, X)$, where $X = \sum_{k=1}^{M} \cos(\vartheta_k)$, $Y = \sum_{k=1}^{M} \sin(\vartheta_k)$, and the atan2 function is a common implementation in many software packages designed to correctly identify the quadrant and magnitude of the angle given by $\tan^{-1}(Y/X)$. The focus about the mean is given by $R = \sqrt{X^2 + Y^2}\,/M$; values of *R* range from 0 to 1, with a value of 1 indicating that all angles contained in a bin are exactly aligned, and lower values of *R* indicating a greater degree of scatter between the binned angles. *R* is very similar to the polarisation order parameter used across multiple studies of collective movement (7, 8) to characterise instantaneous alignment within groups, with the difference being that the polarisation is determined for particular times, whereas the angles used to determine *R* here are aggregated across many different time steps.

S2 Some notes on the accuracy of methods for inferring rules of interaction

The method used in this study is a force-matching/social force-based averaging method (1-3, 9, 10), with the resulting graphs of the functions fitted to the data sometimes referred to as "force maps" (11, 12). This approach has been used across a number of experimental studies, including (1-3, 12-14), with more recent research starting to focus on the accuracy of the method (10, 11). Some of the potential issues with the force-matching approach include the need to better understand which independent variables need to be included in the analysis to accurately estimate interactions (here we use only the relative (x, y) coordinates of neighbours), and the fact that the approach does not have explicit mechanisms for disentangling elements of interactions (11). Such elements may include the drivers for orientation- and attraction-like behaviour, which may be entangled because they apply and are combined over the same spatial range, or because these elements are additive and apply over different spatial ranges, but are combined due to interactions with multiple neighbours (as in the model of (8)). The accuracy of force-matching has been scrutinised in (10) using data from a large number of simulations of the models developed in (8, 15) across a variety of emergent behaviours. As noted in the introduction of the main text, the social force-based approach does seem to be reasonably capable of accurately extracting interactions, specifically those relating to repulsion and attraction, even when data is relatively limited (at about 1000 time steps of data per observation) (10). However, the accuracy of the force maps produced can be affected by group size, which might be reasonably expected based on the way data is aggregated across multiple individuals, and by consistent patterns in group-level movement, such as persistent milling for example (10). Group size effects are unlikely to have played a role in the results examined in the case study of pairs of eastern mosquitofish that forms part of the work here, but should be kept in mind for analyses of larger groups. Provided that comparisons are made between groups of equal sizes though, these effects may not impact the accuracy of the randomisation process. It should be noted though that the analysis in (10) does not examine orientation interactions explicitly. This remains an important avenue for further investigation given that force-matching has no explicit mechanism for separating potentially additive effects like orientation and attraction. In addition, additive effects of repulsion and attraction mechanisms were identified as a potential cause of group-size related inaccuracies in force-matching in (10).

Alternative approaches to the method used for extracting interactions here include the fitting of functions described by equations (11, 12, 16-18), resulting in a data driven model, and the machine learning based approach examined in (19). The machine learning approach has also been validated using simulation data, and has the additional advantage that it identifies the relative weightings that should be applied in determining individual responses to multiple partners (19). In general, there is need for a benchmark study that explicitly compares the accuracy and computational speed of the three classes of interaction analysis to help better understand which may be better in a given context. Based on current evidence all three classes seem to be viable partners for a randomisation method like that examined here. However, this statement is made with the caveat that examination of the methods for estimating interactions is an active area of research, with some of the current focus on the accuracy of force-matching when applied to entangled interactions (as noted above, and in (11)).

S3 Computational considerations and efficient implementation of the randomisation methods

The core calculations for estimating rules of interaction and related quantities are $O\left(N^2\right)$ for a group of N individuals due to pairwise comparison of the coordinates of all individuals. (For N individuals, $N^2 - N$ pairwise comparisons are required, taking into account the non-symmetric relationship in relative coordinates between individuals due to the chosen reference system.) In terms of coding, the most straightforward and naïve approach to implementing the randomisation scheme described in the previous section is to place an existing code for estimating rules of interaction inside an additional outer loop over the number of randomisations to be performed, and to modify the code slightly within the loop so that on each iteration, individuals are randomly assigned to categories, and then the full set of calculations to determine the rules of interaction are applied according to this categorisation (this was our original approach). In doing this, all $N^2 - N$ pairwise comparisons between individuals within a group are repeated for every iteration of the randomisation process, ultimately making the calculations very slow, especially for large groups, or large numbers of randomisations.

We reorganised our calculations to avoid repetition of the $N^2 - N$ pairwise comparisons for every separate randomisation, to make our calculations more efficient. To do this, we first perform a once and for all set of calculations to bin data for all measures of interest, maintaining separate sets of bins for each set of observational data, and for each individual within each set. In our code, written in MATLAB, we bin our data in cell arrays; in the original implementation of our code our bins were identified by two indices (corresponding to ranges of relative x- and y-coordinates), but in the more efficient code we added two more indices to further separate binned data based on observation set, and individual identity. Randomisation is applied by combining pre-binned data for randomly chosen sets of individuals across all observation sets, and then determining averages or other derived quantities from the recombined binned data (as required for the particular measure of interest).

S4 On potential test-statistics when a different method for inferring interactions is coupled with randomisation

An additional technical consideration is the form in which the functions describing an individual's change in velocity in response to its group mates are stored. The functions fitted in this study are stored as matrices, and are rendered as line or surface plots directly from these matrices. The test-statistics examined here are constructed with this matrix-based method for describing the functions in mind. If a machine-learning/artificial neural network approach was used to fit the rules of interaction functions, then it is possible that the function would be stored in terms of the connection weights between the nodes in the network, at least as an intermediate step. To estimate the function in a form compatible with the test-statistics described here, the response of the trained network could be recorded for an appropriately chosen grid of inputs, with the outputs then stored in a matrix, which could then be used immediately for derivation of the test-statistics applied here. Test statistics similar to the mean and maximum absolute difference could also be derived from rules of interaction functions expressed via equations, such as those resulting from the methods in (11, 12, 16, 18). In this case, the mean absolute difference could be calculated via application of an appropriate integral expression, and the maximum may be identifiable via application of fundamental calculus, or the measures could be approximated by evaluating the underlying functions at appropriate mesh-points, which would then lead to the same process for calculating test-statistics applied here.

S5 Case study data and classification of leaders and followers in pairs of female eastern mosquitofish

S5.1 Experiments

Female eastern mosquitofish (n = 80) were collected using hand-nets from Lake Northam, Sydney, NSW, (33°53'07" S; 151°11'35" E). Fish were held in 170 l aquaria and were fed flake food *ad libitum*. Fish were kept for at least three weeks prior to experimentation. A square experimental arena (1.5 m × 1.5 m × 0.2 m) was constructed of opaque white perspex and filled to a depth of 7 cm. In two corners of the arena, diagonally opposite one another, we placed an opaque white holding tube (10 cm diameter). For each trial, we selected two fish of similar size (approximately 1.5 - 2.5 cm) and placed one in each of the holding tubes. To test if a fish's experience with the environment subsequently led this individual to occupy positions at the front (or back) of the group, following an acclimation period of five minutes, we either released one (n = 20 trials), or both fish (n = 20 trials) into the arena. If we had only released one fish into the arena, we allowed this fish to explore the arena for five minutes before we released the second fish. In one of the two treatments, therefore, one fish had explored the environment for longer than the other. For our randomisation calculations described in the main text, we combined all data from both these treatments, irrespective of the experience of the fish with the environment or otherwise. We filmed the trials using a Basler avA1600-65kc camera and recorded using StreamPix (version 5) at 40 fps. Fish were filmed for 6 minutes when both fish had been released into the arena. These films were subsequently converted using VirtualDub (version 1.9.11) and the fish were tracked using Ctrax (version 0.5.4), (20). Any ambiguities in fish identities or other elements of the tracked data were resolved using Ctrax's *fixerrors* GUI.

S5.2 Classification of leaders and followers

Ultimately, we classified fish as ‘leaders’ or ‘followers’ based on an examination of their positions in the pair relative to the group centroid, and correlations in direction of motion as a function of time-lag, when the fish were less than or equal to 100 mm apart. The 100 mm separation threshold corresponded to approximately 4 body lengths, which is a standard scale for determining if shoaling fish are members of the same group, see for example (21-23). 58.49% of our trajectory data for which both fish were moving in the arena satisfied the 100 mm or less separation condition, including the initial periods of time taken for individuals to locate one another.

We first determined the front to back order of the pair of fish relative to the direction of motion of the group centre as follows. For each video frame we identified the mean coordinates of the pair of fish, denoted $\left(\overline{x}(t), \overline{y}(t)\right)$, and treated this as the centre of the pair.

We then estimated the components of velocity of the group centre using standard forward difference approximations (as we did for the components of an individual’s velocity) as follows:

$$u_c(t) = \frac{\overline{x}(t+\Delta t) - \overline{x}(t)}{\Delta t}, \quad v_c(t) = \frac{\overline{y}(t+\Delta t) - \overline{y}(t)}{\Delta t}. \tag{16}$$

For each time step (except for the last where there was no associated estimate for the velocity), we shifted the coordinates of each fish so that the origin of the coordinate system lay at the group centroid by subtracting the coordinates of the group centre from each individual’s coordinates. We then rotated the coordinates of the fish so that the direction of motion of the group centre was parallel to positive *x*-axis using a rotation matrix that rotated

all coordinates through the negative of the angle associated with the velocity given by equations (16). In this transformed coordinate system, we treated the fish with the greatest x-coordinate as being at the front of the pair for a given discrete time. We counted the number of frames that each fish was located at the front of the pair, and retained the relative front or back positions of both group members for all time steps. Tables S1 and S2 detail the number of frames that all pairs were within 100 mm of each other, and the proportion of these frames that each fish occupied the front most position. In general, individuals swapped positions throughout most experiments, even though one individual was more often found at the front of the pair (Figure S2). During our examination of the data, we noted that many instances of occupying the front most position only lasted for short durations. When the front position was in contest, the front most fish relative to the group centroid would often swap multiple times. Figure S3A illustrates the relative frequency that fish spent differing unbroken durations at the front of their pair. The histogram in Figure S3A is dominated by short duration instances of occupying the front. However, the large number of short duration instances of occupying the front position only contributed a small amount to the total duration of data where individuals where pairs were closely grouped, as illustrated in Figure S3B. The shortest duration instances (of duration 0.025 second = 1 frame) only made up 0.99% of the data, whereas instances where fish occupied the front of their pair for more than 1 second (40 frames) corresponded to 86.37% of the overall data. Thus, much of our data was representative of occupancy of the front most position for relatively substantial durations.

Table S1: The total number of frames where a pair of fish were in close proximity to each other (within 100 mm), the proportion of frames spent at the front by each fish, the time lag associated with maximum correlation in direction of motion when fish were either at the front or the back, and the associated maximum correlation in direction of motion as a function of time lag. For these groups the fish identified as fish 1 had an additional 5 minutes to familiarise itself with the tank before fish 2 was released. Bold type indicates the largest proportion of time spent at the front of a pair.

				τ_{ij}^{*} (s)				$C_{ij}\left(\tau_{ij}^{*}\right)$	
Group	No. frames closely grouped	Prop. fish 1 in front	Prop. fish 2 in front	Fish 1 in front	Fish 1 behind	Fish 2 in front	Fish 2 behind	Fish 1 in front (Fish 2 behind)	Fish 2 in front (Fish 1 behind)
1	11655	**0.5887**	0.4113	0.700	-0.550	0.550	-0.700	0.9335	0.9101
2	11690	**0.5377**	0.4623	0.875	-0.875	0.875	-0.875	0.8828	0.8996
3	9844	**0.6790**	0.3210	0.750	-0.875	0.875	-0.750	0.9162	0.8894
4	11645	**0.6670**	0.3330	0.950	-0.650	0.650	-0.950	0.9242	0.8926
5	12671	0.4686	**0.5314**	0.850	-0.675	0.675	-0.850	0.8588	0.8807
6	3704	0.2125	**0.7875**	1.475	-1.100	1.100	-1.475	0.9177	0.7403
7	1340	**0.5657**	0.4343	-3.000	-2.475	2.475	3.000	0.6368	0.9853
8	9917	**0.5301**	0.4699	0.700	-0.675	0.675	-0.700	0.9009	0.8871
9	10120	0.4941	**0.5059**	0.825	-0.900	0.900	-0.825	0.8298	0.8827
10	10402	**0.8101**	0.1899	0.625	-0.875	0.875	-0.625	0.9418	0.8225
11	5019	0.4082	**0.5918**	1.125	-0.400	0.400	-1.125	0.8804	0.8956
12	6565	**0.7555**	0.2445	0.575	-0.875	0.875	-0.575	0.8783	0.7246
13	1147	**0.5004**	0.4996	1.575	-1.575	1.575	-1.575	0.8755	0.7979
14	961	0.3018	**0.6982**	-2.400	-1.150	1.150	2.400	0.9734	0.7829
15	7694	0.4082	**0.5918**	0.675	-0.675	0.675	-0.675	0.9246	0.8983
16	3537	0.4515	**0.5485**	1.050	-0.675	0.675	-1.050	0.8940	0.8306
17	10110	0.3910	**0.6090**	0.700	-0.700	0.700	-0.700	0.8587	0.9127
18	4569	0.4730	**0.5270**	1.200	-1.150	1.150	-1.200	0.7852	0.7880
19	595	0.4723	**0.5277**	1.000	-1.050	1.050	-1.000	0.4320	0.7499
20	9093	0.4576	**0.5424**	0.925	-0.800	0.800	-0.925	0.8497	0.8718

Table S2: The total number of frames where a pair of fish were in close proximity to each other (within 100 mm), the proportion of frames spent at the front by each fish, the time lag associated with maximum correlation in direction of motion when fish were either at the front or the back, and the associated maximum correlation in direction of motion as a function of time lag. Bold type indicates the largest proportion of time spent at the front of a pair.

				τ_{ij}^{*}				$C_{ij}\left(\tau_{ij}^{*}\right)$	
Group	No. frames closely grouped	Prop. fish 1 in front	Prop. fish 2 in front	Fish 1 in front	Fish 1 behind	Fish 2 in front	Fish 2 behind	Fish 1 in front (Fish 2 behind)	Fish 2 in front (Fish 1 behind)
21	8133	**0.6957**	0.3043	1.125	-1.300	1.300	-1.125	0.8492	0.7333
22	6870	0.0357	**0.9643**	2.400	-0.700	0.700	-2.400	0.8634	0.9520
23	8748	**0.8994**	0.1006	0.600	-0.850	0.850	-0.600	0.9613	0.8427
24	9820	**0.7011**	0.2989	0.700	-0.700	0.700	-0.700	0.9223	0.8983
25	10714	0.2976	**0.7024**	1.875	-1.125	1.125	-1.875	0.7638	0.8098
26	4762	0.4794	**0.5206**	0.900	-1.000	1.000	-0.900	0.9382	0.9142
27	3811	0.3836	**0.6164**	0.975	-0.950	0.950	-0.975	0.8836	0.8850
28	5761	**0.6804**	0.3196	0.825	-1.075	1.075	-0.825	0.8932	0.9026
29	5954	0.2538	**0.7462**	0.875	-0.675	0.675	-0.875	0.9183	0.9445
30	9157	0.3374	**0.6626**	0.675	-0.575	0.575	-0.675	0.9197	0.9194
31	6301	0.0689	**0.9311**	-2.650	-0.650	0.650	2.650	0.6307	0.8182
32	7871	**0.9625**	0.0375	0.650	0.875	-0.875	-0.650	0.9093	0.8933
33	5202	0.3754	**0.6246**	-0.900	-1.300	1.300	0.900	0.6792	0.7006
34	3912	**0.5378**	0.4622	1.000	-1.100	1.100	-1.000	0.8085	0.7562
35	3091	**0.7383**	0.2617	0.800	1.725	-1.725	-0.800	0.8166	0.8767
36	242	0.4793	**0.5207**	-1.325	-1.050	1.050	1.325	0.0358	-0.4376
37	2962	**0.7269**	0.2731	0.700	-1.150	1.150	-0.700	0.9152	0.8622
38	2887	0.4288	**0.5712**	0.675	-0.725	0.725	-0.675	0.9015	0.9331
39	7434	0.1987	**0.8013**	0.900	-0.775	0.775	-0.900	0.8406	0.9388
40	6872	**0.8615**	0.1385	0.800	-0.825	0.825	-0.800	0.9386	0.8371

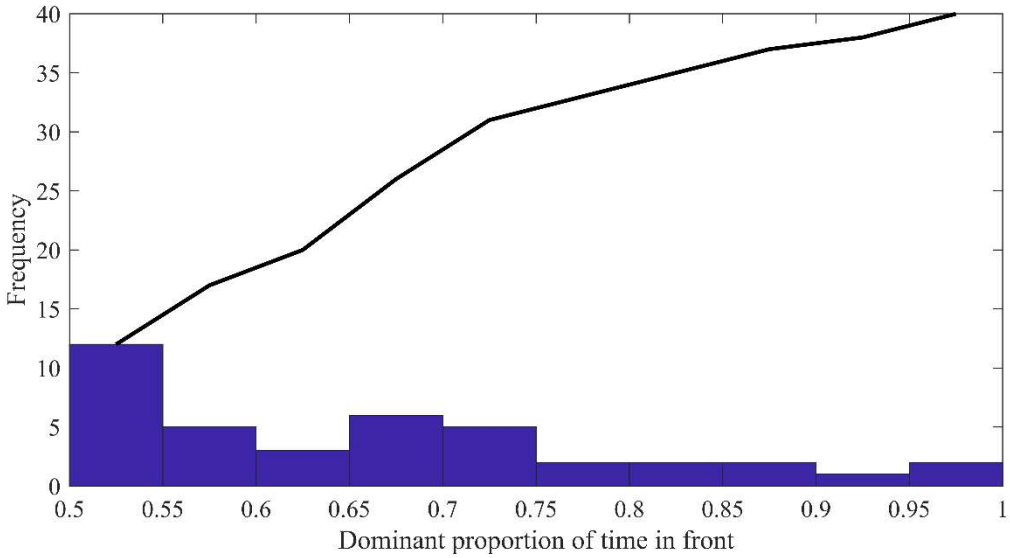

Figure S2: Total proportions of time spent at the front of mosquitofish pairs by those that spent the majority of time at the front (when fish were separated by 100 mm or less). The black line illustrates the cumulative sum of the frequency counts.

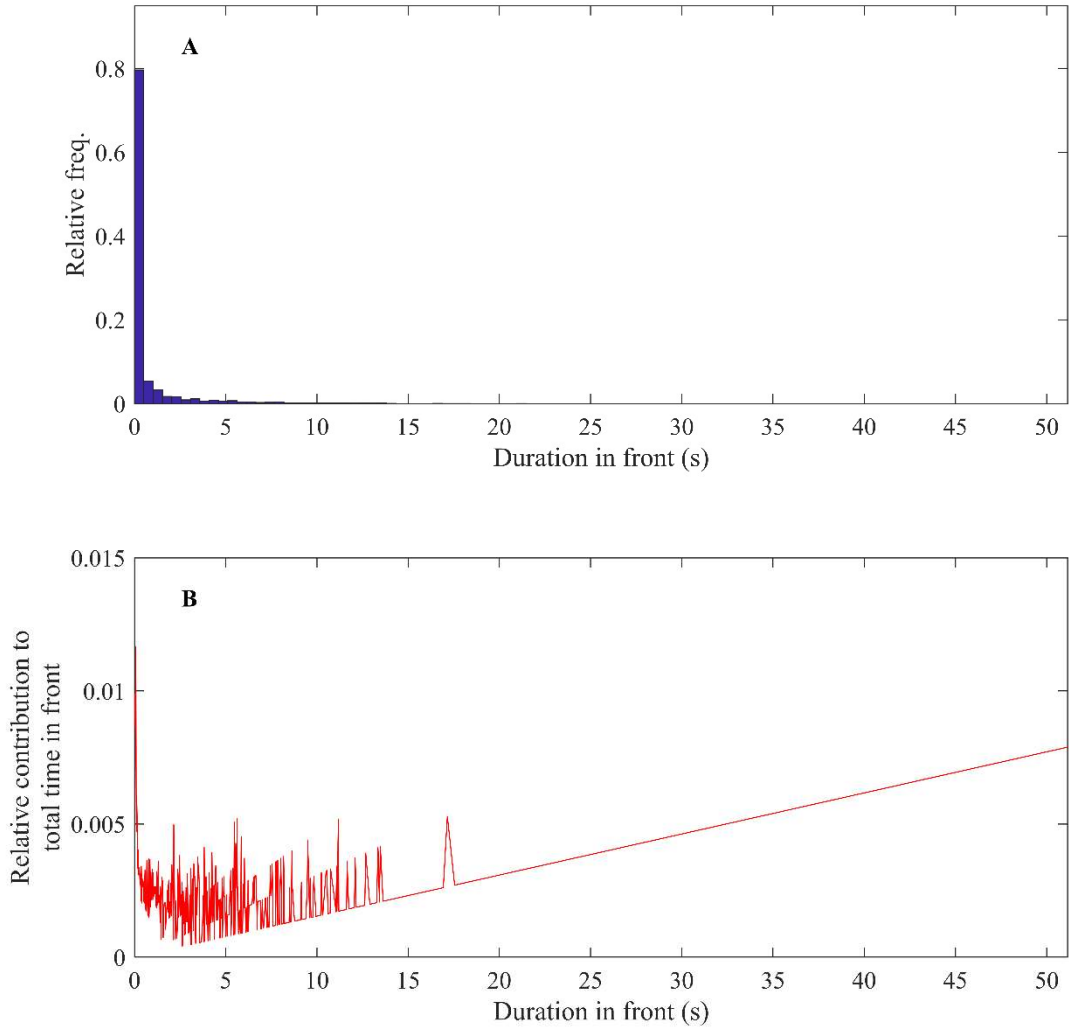

Figure S3: A – the relative frequency of unbroken durations spent at the front of a pair by all fish. **B** – the relative contribution to the total time spent in front by all individuals by separate unbroken durations of front occupancy of different durations.

With reference to our analysis of within pair positioning, we then examined the average directional correlation between the two fish in each pair as a function of a delay time when the fish were either at the front or back of the pair (2, 24). The purpose of this analysis was to determine the relative influence that fish had on the direction of motion of their partner when they were either at the front or back. We identified the fish in each set of observations as fish 1 or 2 consistent with the labelling used in Tables S1 and S2, and then produced two sets of time series of each fish's direction of motion as described in component form by equation (4). The first of these series only retained data for instances where fish 1 was in front (with entries where fish 2 was in front blanked out), and the second series only retained data for instances where fish 2 was in front. Then for each set of time series, we determined the average correlation in directions of motion between the fish

$$C_{ij}(\tau) = \left\langle \hat{\mathbf{u}}_i(t) \bullet \hat{\mathbf{u}}_j(t+\tau) \right\rangle, \tag{17}$$

where $\tau = \tau_k \Delta t$ is time-lag in seconds, $\tau_k \in \{-120, -119, \ldots, 120\}$ was the number of frames corresponding to a given time-lag, and $\langle \cdot \rangle$ represents the mean taken over all t. We then examined the maximum value of $C_{ij}(\tau)$ for each fish in both scenarios (fish 1 or 2 in front), and the corresponding time lag at which this maximum occurred, denoted τ_{ij}^*. Provided that $C_{ij}(\tau_{ij}^*)$ was large enough to suggest reasonable correlation in directions of motion, a positive value for τ_{ij}^* would suggest that fish *j* adjusted its direction of motion to match that adopted by fish *i* at an earlier time (fish *j* was following the direction of fish *i*), whereas negative τ_{ij}^* would suggest that fish *i* was following fish *j*. Tables S1 and S2 also detail τ_{ij}^* and the corresponding maximum correlation for all 40 pairs of fish, and Figures S4 to S7 show plots of $C_{ij}(\tau)$ for each pair of fish. Based on a direct examination of τ_{ij}^* and $C_{ij}(\tau_{ij}^*)$, as listed in Tables S1 and S2, in the majority of cases (for 73 out of 80 fish), when a fish occupied the front of a pair, then its partner would follow the direction of motion of the front fish. Similarly, 73 out of 80 fish would tend to follow the direction of motion of their partner when the focal fish occupied the back of the pair. The delay time associated with the maximum correlation for directions of movement to flow from the back of the pair (τ_{ij}^*) was usually between 0.5 and 1.5 seconds.

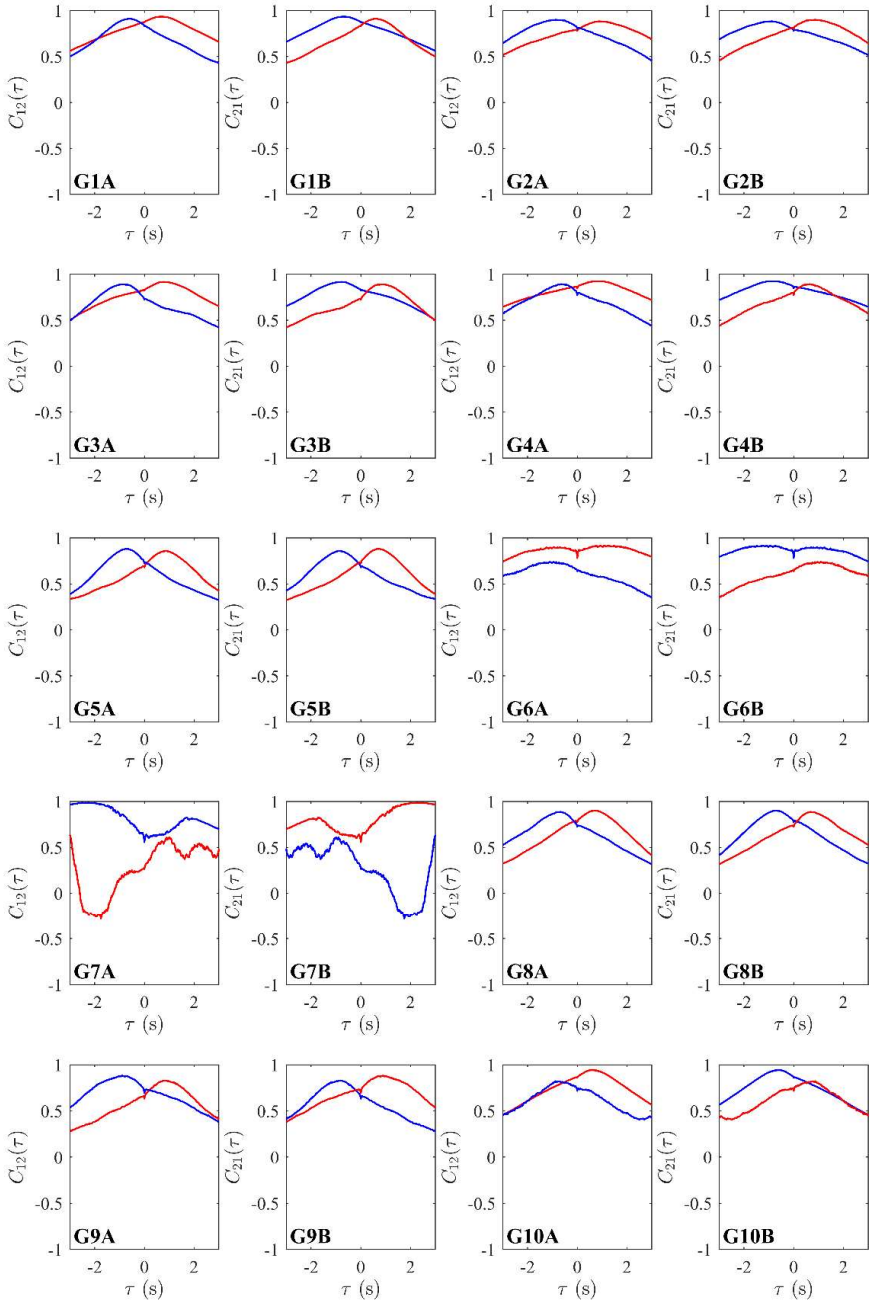

Figure S4: Directional correlations as a function of time-lag for groups 1 to 10 when focal individuals occupied the front of pairs (red curves) and when focal individuals where at the back of pairs (blue curves). Panels with identifiers ending in A treat fish 1as the focal individual, and panels with identifiers ending in B treat fish 2 as the focal individual.

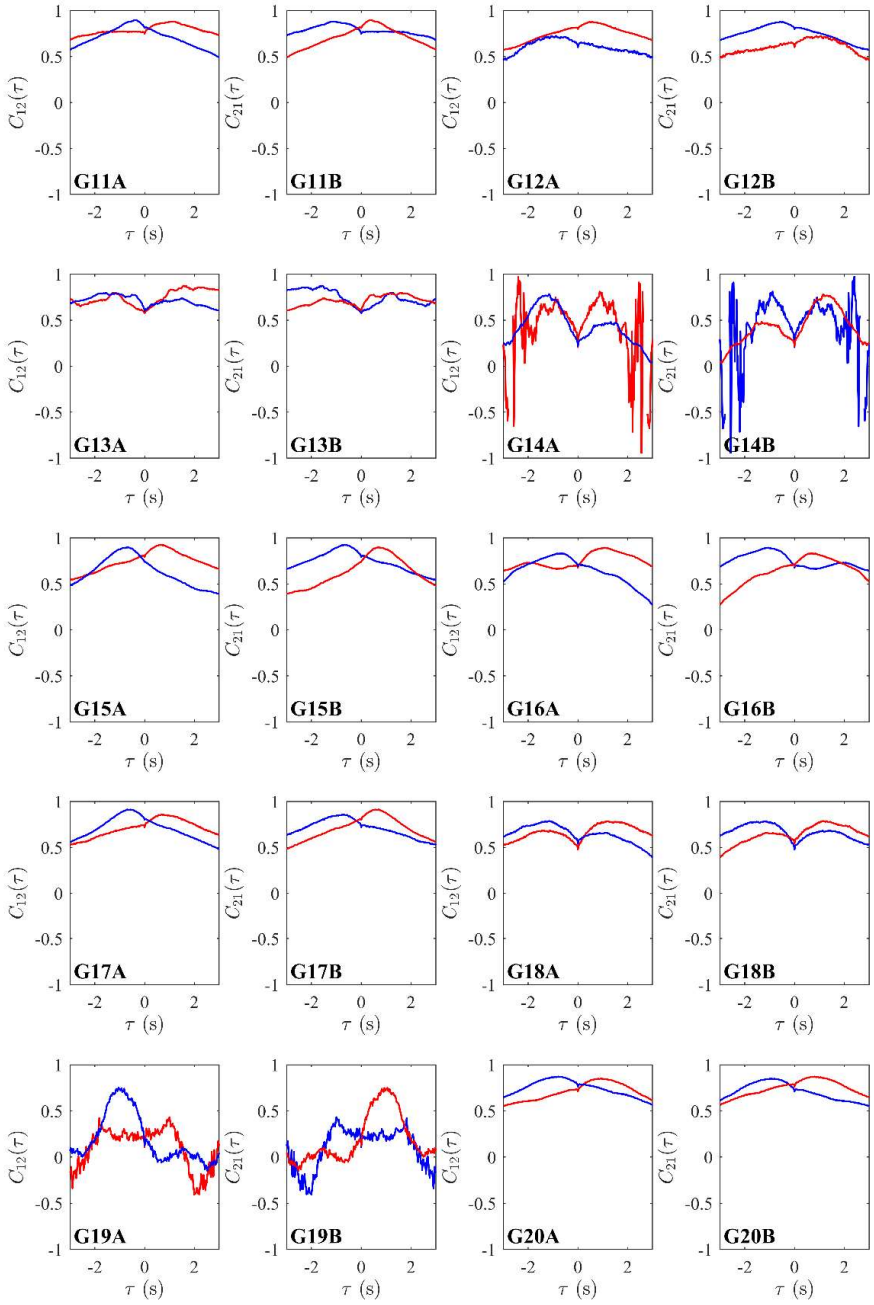

Figure S5: Directional correlations as a function of time-lag for groups 11 to 20 when focal individuals occupied the front of pairs (red curves) and when focal individuals where at the back of pairs (blue curves). Panels with identifiers ending in A treat fish 1as the focal individual, and panels with identifiers ending in B treat fish 2 as the focal individual.

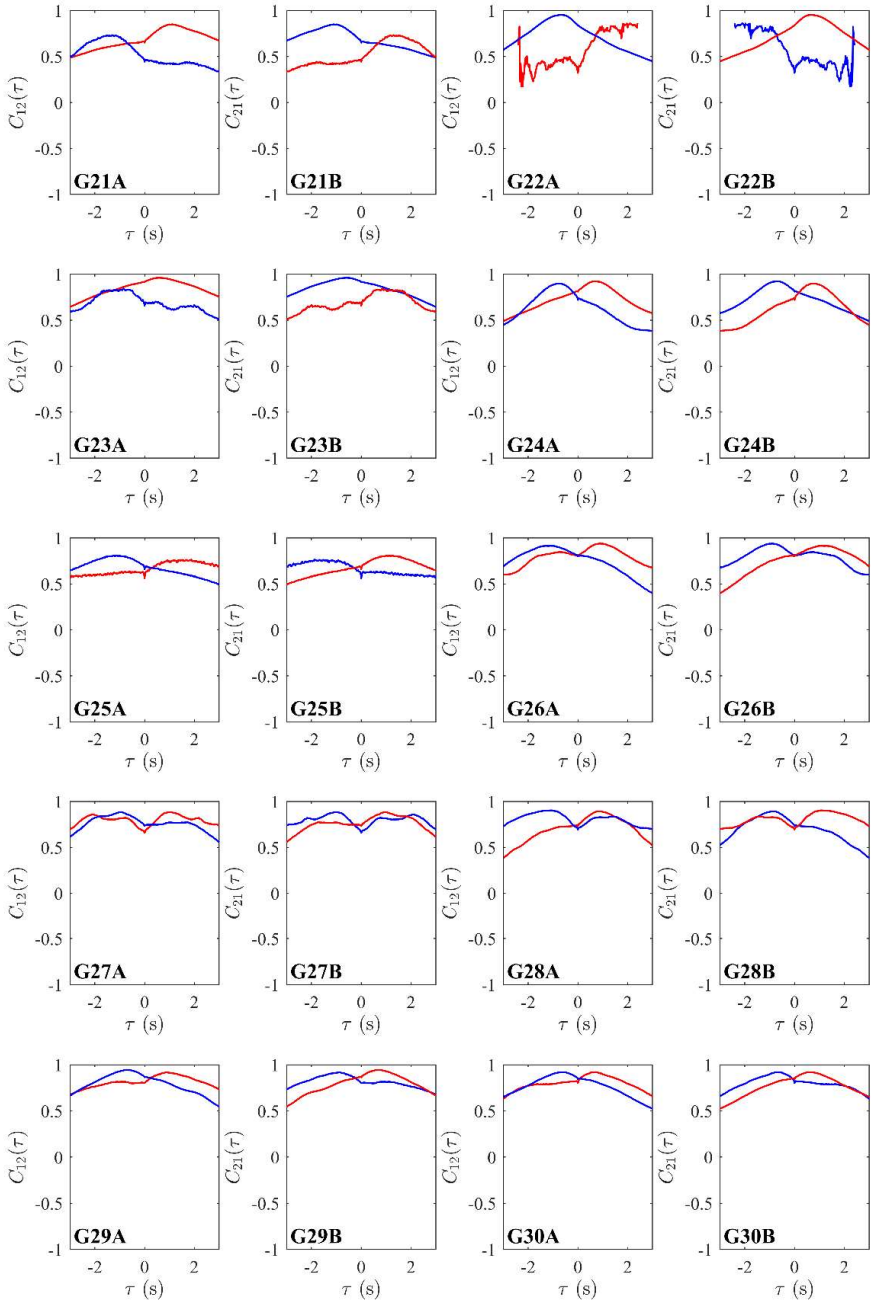

Figure S6: Directional correlations as a function of time-lag for groups 21 to 30 when focal individuals occupied the front of pairs (red curves) and when focal individuals where at the back of pairs (blue curves). Panels with identifiers ending in A treat fish 1as the focal individual, and panels with identifiers ending in B treat fish 2 as the focal individual.

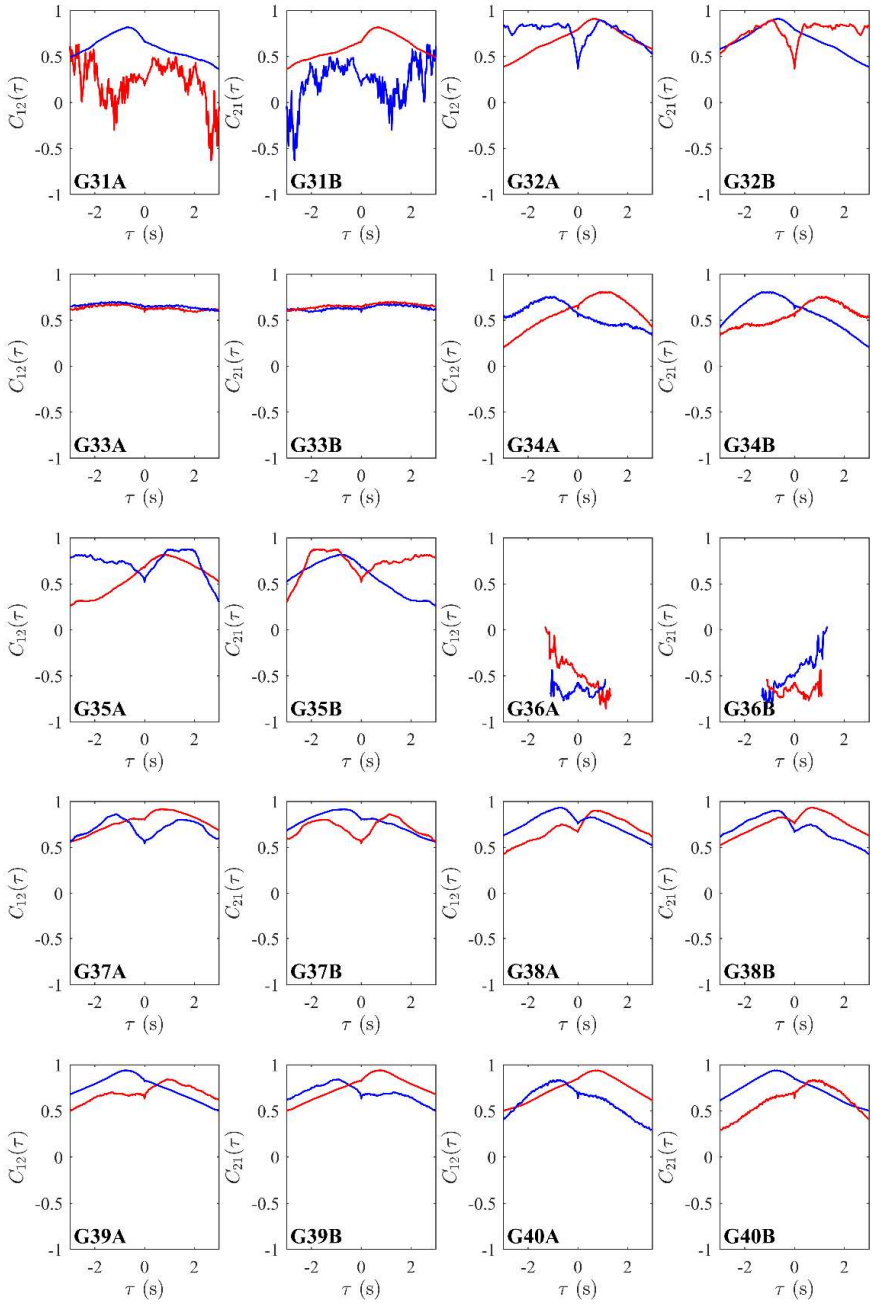

Figure S7: Directional correlations as a function of time-lag for groups 31 to 40 when focal individuals occupied the front of pairs (red curves) and when focal individuals where at the back of pairs (blue curves). Panels with identifiers ending in A treat fish 1as the focal individual, and panels with identifiers ending in B treat fish 2 as the focal individual.

Examination of the graphs in Figures S4 to S7 reveals that although in many cases C_{ij} is represented by a relatively simple curve with a single peak which is more easily interpretable in terms of leader/follower relationships, in some cases the curves for C_{ij} are more complex, with multiple peaks (as is most evident for curves for groups 6, 7, 13, 14, 18, 19, 27, 28, 35, and 37), relatively large amounts of noise (for groups 14, 19, 22, 31, and 36), or that are relatively flat (group 33). Groups 22 and 31 were cases where one individual heavily dominated the front of the pair, with the fish identified as fish 2 in these pairs occupying the front position for 96.43% and 93.11% of the data respectively (corresponding to the smooth curves). Thus relatively little data was used to generate the curves corresponding to instances where the other fish (fish 1) occupied the front position, which may have led to the noisy appearance of these curves. Groups 14, 19, and 36 were closely grouped for less than 1000 frames each, which may again have resulted in the noisy looking curves for C_{ij}.

In spite of the complexities discussed above, overall, the analysis suggests that the majority of the fish tended to lead movement directions when at the front of a pair. Thus, if a mosquitofish was at the front of their pair for the greatest proportion of time when closely grouped, then that fish would most often lead the movement direction of the pair. Hence, for each pair, we categorised the fish that occupied the front most position for the greatest proportion of time as “leaders”, and their partners as “followers”.

S6 Supplementary results

Baseline functions associated with leaders and followers for measures of individual speeds and relative alignment are presented in Figures S8 to S12, along with summaries of randomisation calculations.

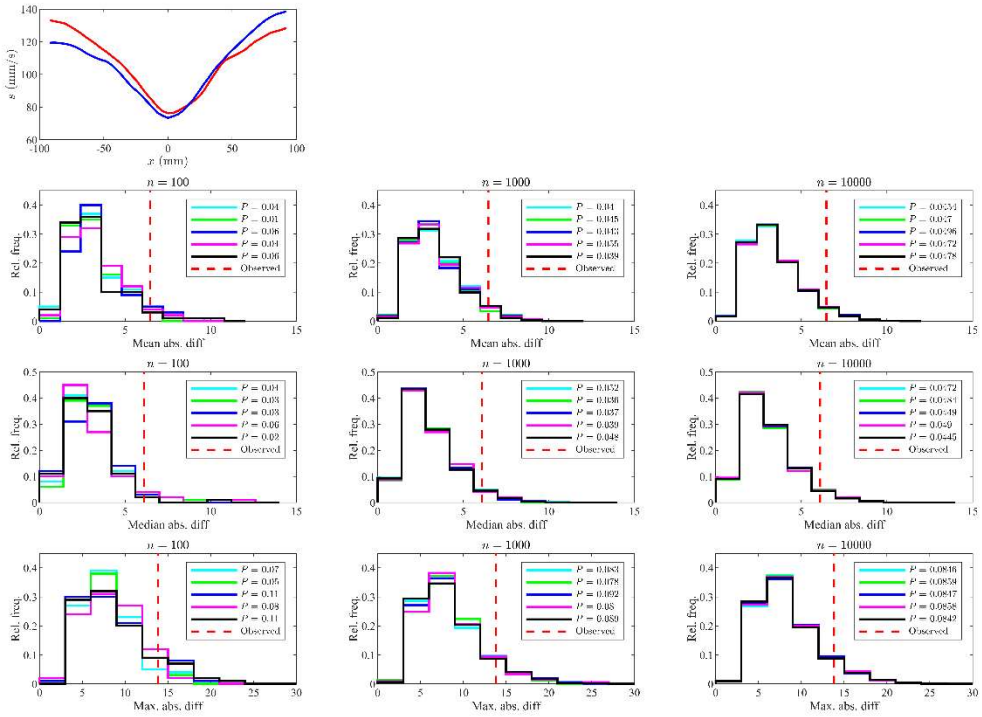

Figure S8: Top left panel: the mean speed, s, of leaders (red curve) and followers (blue curve) as a function of the relative x-coordinate of their partner. In this plot the focal individual is located at the origin, and is travelling parallel to the positive x-axis (from left to right). Lower panels summarise the results of the randomisation analysis, following the structure used in Figures 1 to 9.

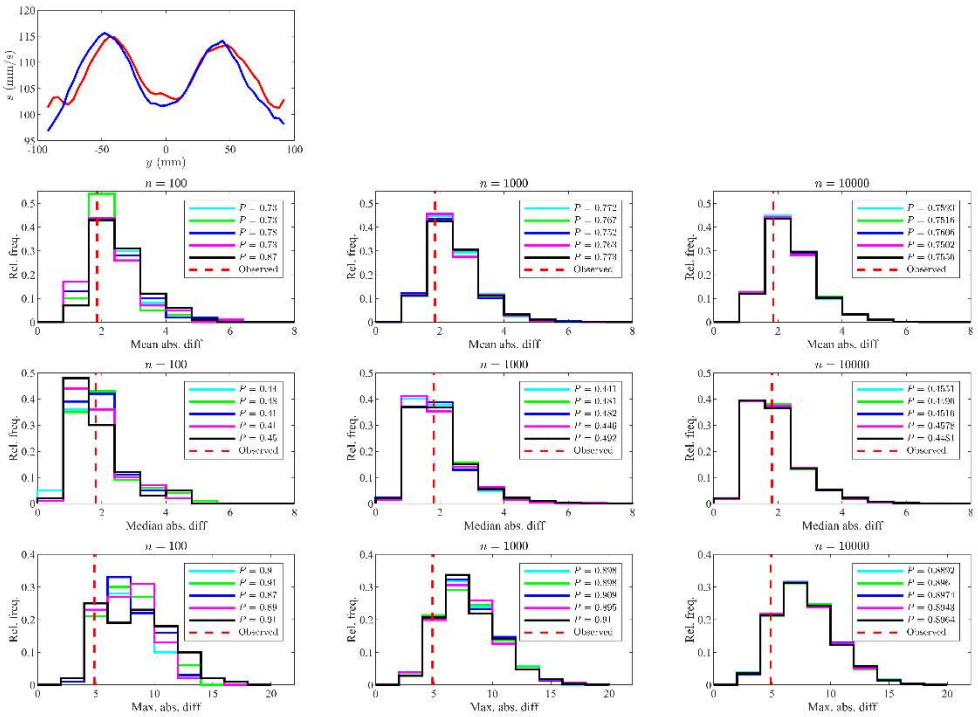

Figure S9: Top left panel: the mean speed, s, of leaders (red curve) and followers (blue curve) as a function of the relative y-coordinate of their partner. In this plot the focal individual is located at the origin, and is travelling parallel to the positive x-axis (out of the page, and towards the reader). Lower panels summarise the results of the randomisation analysis, following the structure used in Figures 1 to 9.

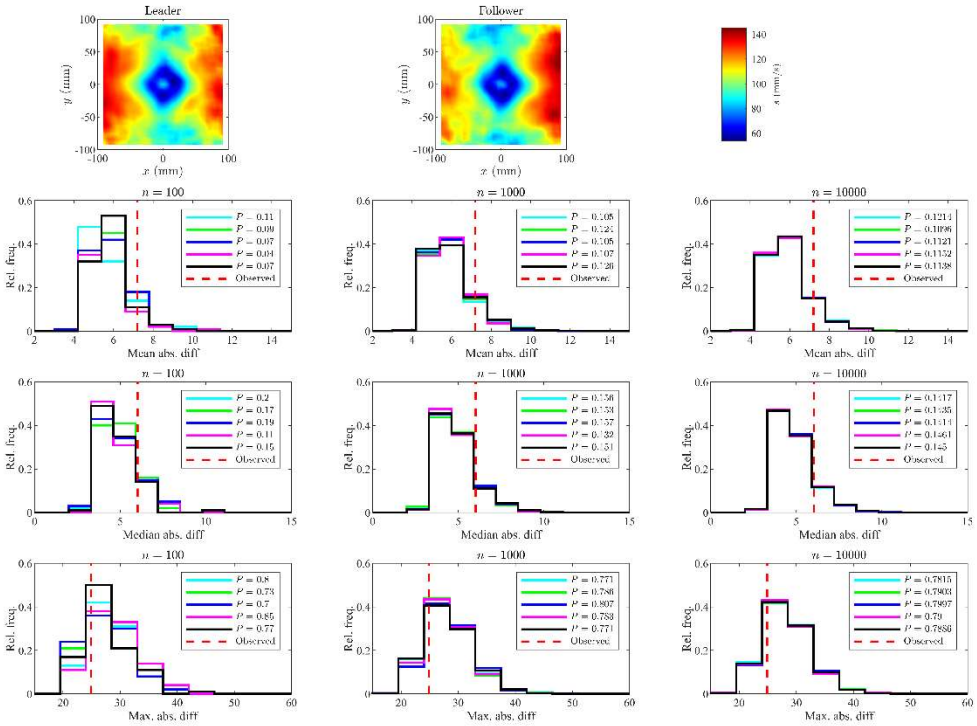

Figure S10: Top row: the mean speed, s, of leaders (left), and followers (centre) as a function of the relative (x, y) coordinates of their partners. In these plots the focal individual is located at the origin, and is travelling parallel to the positive x-axis (from left to right). The colour scale on these plots is such that bluer regions correspond to lower (but non-zero) speeds, and redder regions correspond to greater speeds (right colour bar). Lower panels summarise the results of the randomisation analysis, following the structure used in Figures 1 to 9.

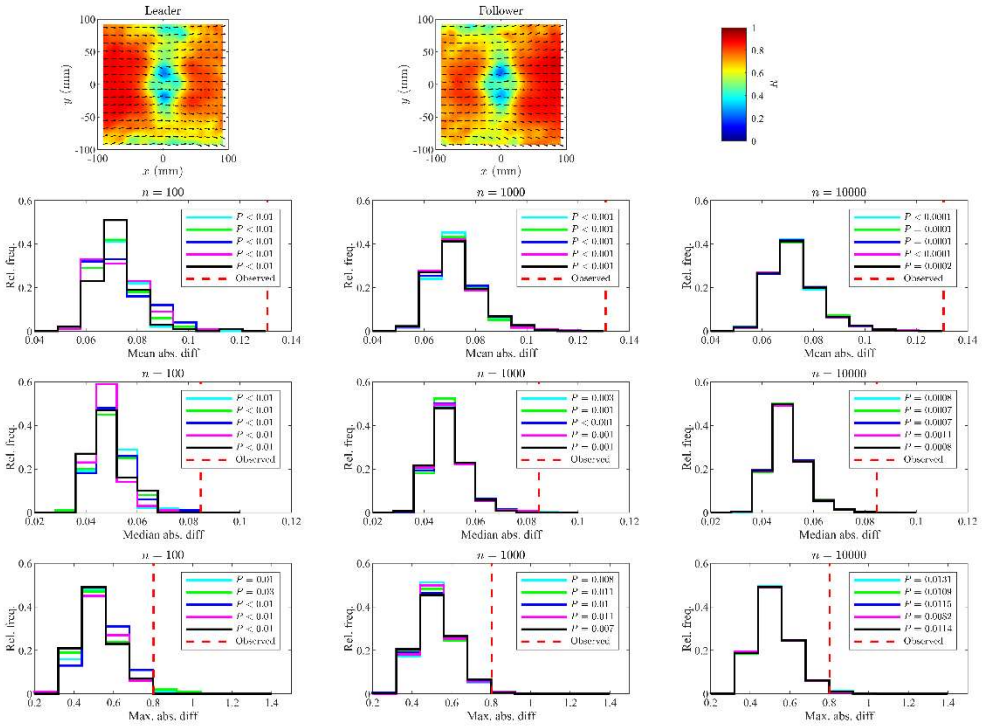

Figure S11: Top row: mean directions of motion, $\overline{\vartheta}$, (arrows) and the focus about these mean directions, R, (colours) of partner fish to leaders (left) and followers (centre) as a function of the relative (x, y) coordinates of partner fish. In these plots the focal individual is located at the origin, and is travelling parallel to the positive x-axis (from left to right). Redder regions on these graphs correspond to greater focus/less variability of binned angles about the mean direction, whereas bluer regions indicate lesser focus/greater variability. The lower rows summarise the results of our randomisation calculations as applied to the mean relative directions of partner fish as a function of (x, y).

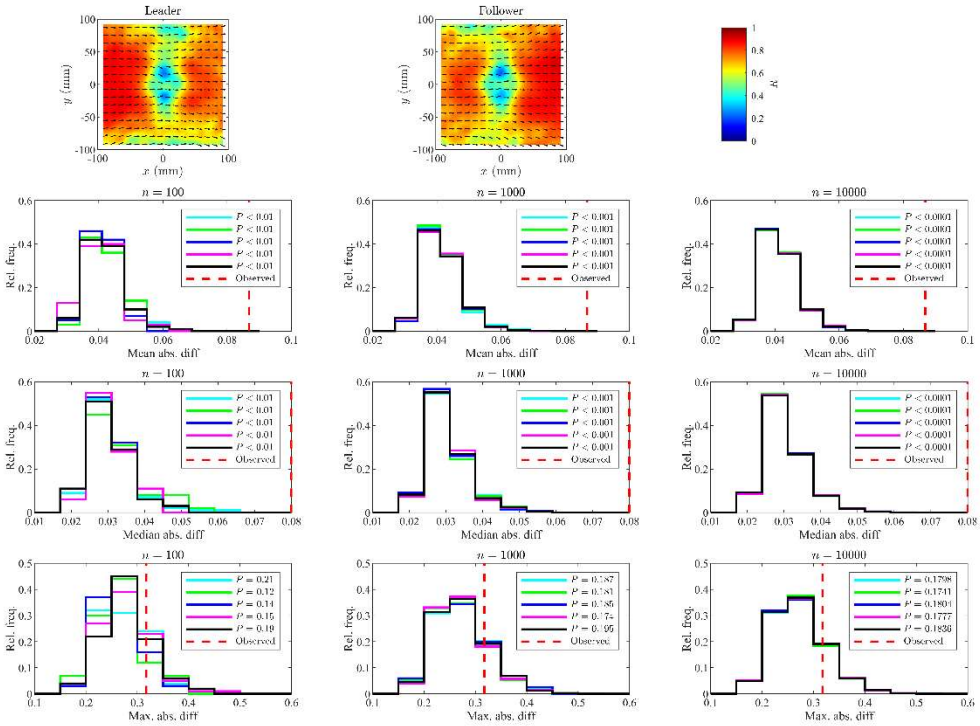

Figure S12: Top row (a duplication of the top row in Figure S11): mean directions of motion, $\overline{\vartheta}$, (arrows) and the focus about these mean directions, R, (colours) of partner fish to leaders (left) and followers (centre) as a function of the relative (x, y) coordinates of partner fish. In these plots the focal individual is located at the origin, and is travelling parallel to the positive x-axis (from left to right). Redder regions on these graphs correspond to greater focus/less variability of binned angles about the mean direction, whereas bluer regions indicate lesser focus/greater variability. The lower rows summarise the results of our randomisation calculations as applied to the variability measure $R(x, y)$.

S6.1 Statistical considerations

The randomisation tests identified significant differences between leaders and followers for nine measures across all five sets of randomisations with 100, 1000, and 10000 iterations. In addition to the seven measures detailed in the main text, differences were consistently identified at all numbers of randomisations for $\overline{\vartheta}$ as a function of (x, y), and R as a function of (x, y) (Table S3). Further to the case where not all repeat calculations of the maximum absolute difference test identified significant differences in $\Delta s/\Delta t$ as a function of x when $n = 100$ or $n = 1000$ (discussed in the main text), ambiguities also arose in the significance (or not) of differences in s as a function of x for all three randomisation tests and s as a function of (x, y) for the mean absolute difference test (Table S4). These ambiguities resolved when n was increased to 10000.

Table S3: Additional measures for which significant differences between leaders and followers were identified for all 5 repeat randomisation tests with $n = 100$, $n = 1000$, and $n = 10000$ for given test-statistics. Here $\overline{\vartheta}$ is the mean direction of motion of a partner as a function of their relative coordinates, and R is a measure of the variability of relative partner directions at given relative coordinates.

Measure	Test-statistic(s)
$\overline{\vartheta}$ as a function of (x, y)	mean absolute difference, median absolute difference, maximum absolute difference
R as a function of (x, y)	mean absolute difference, median absolute difference

Table S4: Cases where some, but not all, randomisation tests identified significant differences in given measures between leaders and followers for a given test-statistic. Numerical values in the rightmost three columns are the number of repeat tests out of 5 for which a significant difference (with $P < 0.05$) was identified. Numbers in parentheses include all cases where $P \leq 0.05$. Here s is the mean speed of individuals as a function of the relative coordinates of their partner.

Measure	Test-statistic	$n = 100$	$n = 1000$	$n = 10000$
s as a function of x	mean absolute difference	3	4	5
s as a function of x	median absolute difference	4	4	5
s as a function of x	maximum absolute difference	0 (1)	0	0
s as a function of (x, y)	mean absolute difference	1	0	0

Table S5 tabulates the total number of instances where distributions of test-statistics generated via randomisation were identified as being significantly different by two-sample Kolmogorov-Smirnov tests with Holm-Bonferroni corrections to significance for multiple pairwise comparisons. There were no differences in distributions of randomised test statistics under the maximum absolute difference test applied to any measure for any of the tested values of n.

Table S5: Number of instances where pairs of test-statistic distributions generated via randomisation were identified as significantly different by two-sample Kolmogorov-Smirnov tests with Holm-Bonferroni corrections to significance for multiple pairwise comparisons.

Measure	Test-statistic	n	Number of pairs (out of 10)
s as a function of y	mean absolute difference	100	1
$\frac{\Delta\theta}{\Delta t}$ as a function of x	median absolute difference	100	2
p as a function of (x, y)	median absolute difference	1000	1
$\frac{\Delta\theta}{\Delta t}$ as a function of x	median absolute difference	10000	3

In terms of clearly significant or non-significant differences, exceptions to this form of self-consistency occurred in 6 to 8 (depending on the interpretation of significance when $P = 0.05$) out of the 126 sets of randomisations that we preformed (with these sets identified by the interaction based measure of interest, the test-statistic used, and the number of randomisation iterations). However, such ambiguities were resolved when the number of randomisation iterations was 10000; our work here seems to suggest that when one of the tests generates a P-value close to the threshold for significance, it may be better to apply a larger number of randomisation iterations to verify the result. In terms of calculation time, with our current code it took approximately one day to complete sets of 10000 randomisations for all 14 measures explored here for one test-statistic (that is, approximately 140000 randomisation iterations in total per day). Even rarer in relative terms were inconsistencies in the form of statistically significant differences between randomisation generated distributions of test-statistics, with 7 pairwise differences identified across 1260 comparisons of these distributions (derived from 10 pairwise comparisons for each interaction measure, test-statistic, and number of randomisation iterations). Six of these differences occurred in tests based on the median absolute difference test-statistic in cases where 100, 1000, and 10000 randomisation iterations were applied. Only one difference was identified between a pair of randomisation generated test-statistic distributions for the mean absolute difference test, for 100 randomisation iterations, and there were no such differences identified under the maximum absolute difference tests. In addition, visual inspection of the distributions of randomisation derived test-statistics suggests convergence of distributions under repeated tests as the number of randomisations increases, consistent with the results of our pairwise comparisons via Kolmogorov-Smirnov tests. Based on our results, all three tests seem very reliable in terms of self-consistency and repeatability.

S6.2 Further insights into leader follower pairs of eastern mosquitofish

General behaviour

The mosquitofish moved at their lowest speeds on average when their partners occupied the region extending out to about 30 to 40 mm from the location of focal individuals; this is a similar region to where individuals moderated their speed consistent with collision avoidance (Figures S8, S9, and S10). The fish also adopted lower mean speeds when their partners were to their side, at distances of 80 to 100 mm, and tended to move faster when their partners were beyond about 50 mm in front or behind. To our knowledge, measurements of the mean speed of mosquitofish as a function of partner location of the type detailed here have not been reported before, but similar tendencies to move at greater speeds when partners are at some

distance to the front and back have been observed for three-spine sticklebacks (4) and x-ray tetras (3) (derived using the methods described in this paper). However, both those species tended to move slowest when partners were close, and to the front or back, but moved at moderate or faster speeds when their partners were close and to the sides, whereas the mosquitofish examined here moved slower when partners were close, irrespective of if they were located along the front-back axis, or the side.

On average, the fish were quite well aligned in movement direction, particularly when partners occupied the approximate strip where $-100 < x < 100$ mm, $-25 < y < 25$ mm (Figures S11 and S12). There was less variation in relative directions of motion when partners occupied regions with $x < -30$ mm or $x > 30$ mm (as evidenced by relatively high values of R). The greatest variation in relative directions of motion occurred when partners occupied small approximately circular regions to the left and right of focal individuals at close range (as evidenced by low values of R, and visualised as small blue regions in Figures S11 and S12).

Significant differences

The mean and median absolute difference tests suggested that there were significant differences in the speed of leaders and followers as a function of the relative x-coordinates of their partners. Leaders tended to travel at greater speeds than followers when their partners were behind them, and followers tended to travel at greater speed than leaders when their partners were in front of them (Figure S8).

All three tests suggested that there were significant differences in the relative alignment between leaders and followers as a function of the relative (x, y) coordinates of partners, however it is difficult to discern the details of these differences from the arrow plots in Figures S11 and S12. In addition, the mean and median absolute difference tests suggested significant differences in R as a function of (x, y) between leaders and followers. There tended to be less variation in instantaneous alignment when followers occupied the rear of the pair compared to leaders (as characterised by the highest values of R in Figures S11 and S12).

References

1. Herbert-Read JE, Perna A, Mann RP, Schaerf TM, Sumpter DJT, Ward AJW. Inferring the rules of interaction of shoaling fish. Proceedings of the National Academy of Sciences of the United States of America. 2011;108:18726-31.
2. Katz Y, Tunstrøm K, Ioannou CC, Huepe C, Couzin ID. Inferring the structure and dynamics of interactions in schooling fish. Proceedings of the National Academy of Sciences of the United States of America. 2011;108(46):18720-5.
3. Schaerf TM, Dillingham PW, Ward AJW. The effects of external cues on individual and collective behavior of shoaling fish. Science Advances. 2017;3:e1603201.
4. Ward AJW, Schaerf TM, Herbert-Read JE, Morrell L, Sumpter DJT, Webster MM. Local interactions and global properties of wild, free-ranging stickleback shoals. Royal Society Open Science. 2017;4:170043.
5. Lukeman R, Li YX, Edelstein-Keshet L. Inferring individual rules from collective behavior. Proceedings of the National Academy of Sciences of the United States of America. 2010;107(28):12576-80.
6. Zar JH. Biostatistical Analysis. 3 ed. Upper Saddle River, NJ: Prentice Hall; 1996.
7. Vicsek T, Czirók A, Ben-Jacob E, Cohen I, Shochet O. Novel type of phase-transition in a system of self-driven particles. Physical Review Letters. 1995;75(6):1226-9.
8. Couzin ID, Krause J, James R, Ruxton GD, Franks NR. Collective memory and spatial sorting in animal groups. Journal of Theoretical Biology. 2002;218(1):1-11.
9. Eriksson A, Jacobi MN, Nyström J, Tunstrøm K. Determining interaction rules in animal swarms. Behavioral Ecology. 2010;21(5):1106-11.
10. Mudaliar RK, Schaerf TM. Examination of an averaging method for estimating repulsion and attraction interactions in moving groups. PLoS ONE. 2020;15(12):e0243631.
11. Escobedo R, Lecheval V, Papaspyros V, Bonnet F, Mondada F, Sire C, et al. A data-driven method for reconstructing and modelling socual interactions in moving animal groups. Philosophical Transactions of the Royal Society B. 2020;375:20190380.
12. Zienkiewicz AK, Ladu F, Barton DAW, Porfiri M, Bernardo MD. Data-driven modelling of social forces and collective behaviour in zebrafish. Journal of Theoretical Biology. 2018;443:39-51.
13. Herbert-Read JE, Rosén E, Szorkovszky A, Ioannou CC, Rogell B, Perna A, et al. How predation shapes the social interaction rules of shoaling fish. Proceedings of the Royal Society B. 2017;284:20171126.
14. Ward AJW, Schaerf TM, Burns ALJ, Lizier JT, Crosato E, Prokopenko M, et al. Cohesion, order and information flow in the collective motion of mixed-species shoals. Royal Society Open Science. 2018;5:181132.
15. D'Orsogna MR, Chuang YL, Bertozzi AL, Chayes LS. Self-propelled particles with soft-core interactions: patterns, stability, and collapse. Physical Review Letters. 2006;96:104302.
16. Gautrais J, Ginelli F, Fournier R, Blanco S, Soria M, Chate H, et al. Deciphering interactions in moving animal groups. PloS Computational Biology. 2012;8(9).
17. Zienkiewicz A, Barton DAW, Porfiri M, Bernardo Md. Data-driven stochastic modelling of zebrafish locomotion. Journal of Mathematical Biology. 2015;71:1081-105.
18. Calovi DS, Litchinko A, Lecheval V, Lopez U, Pérez Escudero A, Chaté H, et al. Disentangling and modeling interactions in fish with burst-and-coast swimming reveal distinct alignment and attraction behaviors. PLoS Computational Biology. 2018;14(1):e1005933.

19. Heras FJH, Romero-Ferrero F, Hinz RC, de Polavieja GG. Deep attention networks reveal the rules of collective motion in zebrafish. PLoS Computational Biology. 2019;15(9):e1007354.
20. Branson K, Robie AA, Bender J, Perona P, Dickinson MH. High-throughput ethomics in large groups of Drosophila. Nature Methods. 2009;6(6):451-7.
21. Pitcher T, Magurran AE, Allan JR. Shifts of behaviour with shoal size in cyprinids. Proceedings of the 3rd British Freshwater Fisheries Conference. 1983;3:220-1.
22. Pitcher TJ, Parrish JK. Functions of shoaling behaviour in teleosts. In: Pitcher TJ, editor. Behaviour of Teleost Fishes1993. p. 363-439.
23. Hoare DJ, Couzin ID, Godin JGJ, Krause J. Context-dependent group size choice in fish. Animal Behaviour. 2004;67:155-64.
24. Nagy M, Àkos Z, Biro D, Vicsek T. Hierarchical group dynamics in pigeon flocks. Nature. 2010;464(7290):890-U99.