Electronic Supplementary Material

Valentin Lecheval Li Jiang Pierre Tichit Clément Sire
Charlotte K. Hemelrijk Guy Theraulaz

1 Material and methods

1.1 Experimental procedures and data collection

Fish were purchased from Amazonie Labège (http://www.amazonie.com) in Toulouse, France. They were kept in 150 L aquariums on a 12:12 hour, dark:light photoperiod, at 27.5° C (±0.8° C) and were fed *ad libitum* with fish flakes. The body length of the fish in the experiments was on average 3.4 cm (± 0.44 cm).

Table S1: Number of trials, total duration of trials, number of collective U-turns and average body length of individuals for each group size.

Group Size	Number of trials	Total duration	Total number of collective U-turns	Body length (mm, mean ± se)
1	4	260 min	1058	33.1 ± 1.8
2	10	652 min	1135	33.3 ± 0.8
4	10	684 min	1868	36.1 ± 0.6
5	10	543 min	500	31.5 ± 0.3
8	9	602 min	459	35.9 ± 0.6
10	14	832 min	49	33.4 ± 0.4
20	11	703 min	30	Not available

The experimental tank (120×120 cm) was made of glass and was set on top of a box to reduce vibrations. It was surrounded by four opaque white curtains and illuminated homogeneously by four LED light panels. Inside an experimental tank, a ring-shaped corridor was filled with 7 cm of water of controlled quality (50% of water purified by reverse osmosis and 50% of water treated by activated carbon) heated at 27.6° C (±0.9° C) (Figure S1A). The corridor was 10 cm wide with a circular outer wall of radius 35 cm. The shape of the circular inner wall was conic and its radius at the bottom was 25 cm. The conic shape was chosen to avoid the occlusion on videos of fish swimming too close to the inner wall. Fish were introduced in and acclimatised to the experimental tank during a period of 10 minutes before the trial starts. During each trial of one hour, individuals were swimming freely without external perturbation. Note that six experiments with a single fish have been discarded because of the inactivity of the individuals.

1.2 Data extraction and pre-processing

Except for the group size of 20 fish, the positions of fish on each frame were tracked with idTracker 2.1 [1]. Sometimes, fish were misidentified by the tracking software, for instance when two fish were swimming too close to each other. All sequences that were missing a maximum of 50 consecutive positions were interpolated. For groups of 20 fish, only the number of collective U-turns, the time interval between two consecutive U-turns and the time of individual U-turns during collective U-turns have been recorded manually with VLC media player 2.2.8 and its extension Jump to time (Previous frame) 2.1. For this groups size, one U-turn in which fish almost all performed 3 U-turns resulting in a very long event (14.14 s, while others last for 3.6 s on average) has been discarded from the analysis at the individual scale.

Time series of positions were converted from pixels to meters and the origin of the coordinate system was set to the centre of the ring-shaped tank. Body lengths and headings of fish were measured on each frame using the first axis of a principal component analysis of the fish shape issued by idTracker. Table S1 summarises the data collected in our study.

1.3 Detection and quantification of individual and collective U-turns

When a group of fish makes a collective U-turn, the degree of alignment to the wall averaged over all individuals of the group $\bar{a}(t)$ changes sign. We used this as the criterion for detecting collective U-turns automatically from the smoothed time-series of $\bar{a}(t)$, using a centred moving average over 9 consecutive frames. A collective U-turn in a group of n fish starts when the degree of alignment to the wall $a_i(t)$ of the fish i that initiates the U-turn is 0 and it ends when the degree of alignment to the wall $a_j(t)$ of the last fish j that turns is 0. For each collective U-turn, we ranked the order with which each individual turned r_i (where $r_i = 1$ refers to the individual i initiating it) and the spatial positions of each individual at the initiation of the U-turn. In order to compare the spatial positions of individuals swimming in groups of various shapes, we compute at the beginning of the U-turns $\Phi_i = -(\theta_i - \theta_f)/(\theta_f - \theta_l)$, where the angle $\theta_i - \theta_f$ between each individual and the fish in front of the group, normalised by the angle $\theta_f - \theta_l$ between the first and last fish. We discretised $\Phi \in [0, 1]$ in n cells with increasing indices and the spatial position π_i is given by the index of the cell that contains Φ_i. π_i is 1 if an individual is very close to the front of the group when the first individual turns and n if it is close to the back of the group at this time.

To compute the ranks of turning and the spatial positions of individuals at the initiation of the U-turns, we needed to make sure that fish were responding to the initiation of a specific U-turn (and not to a previous U-turn very close in time). Therefore, we only considered situations where fish were swimming for at least 2 seconds in the same direction before and after the U-turns.

Failed collective U-turns (i.e. U-turns initiated by one or more individuals that are not fully propagated) are also detected. A failed U-turn is detected when the average of the sign of the degree of alignment is not $|1|$ and when the sign of the average degree of alignment does not switch. To address possible noise in experimental data, the average of the sign of the degree of alignment has to be different from $|1|$ during at least 25 frames (half a second).

For a given group size, we compute the average rate of U-turns (failed or not) initiated per individual as

$$\frac{u_n + f_n}{nT_n}, \tag{S1}$$

with n the group size, u_n the number of collective U-turns (fully propagated), f_n the number of failed collective U-turns and T_n the duration of the experiments. The probability that a collective U-turn is fully propagated is computed by

$$\frac{u_n}{u_n + f_n}. \tag{S2}$$

1.4 Data scaling

Data scaling shown in Figure 1 is obtained by finding the value of the time parameter t_n that minimises the least-square error between the normalised degree of alignment with the wall averaged over the U-turns at a given group size n and that averaged over the U-turns of a group size of reference (namely, groups of 5 fish). To compute error bars, t_n has been bootstrapped by applying the least square method randomising the collective U-turns considered in the averaged normalised degree of alignment for each group size. For each group size, $N = 1000$ bootstrapped samples have been obtained. The same method has been used in Figure 4.

1.5 Statistical tests

We used R [2] and the package `lme4` [3] to perform a linear mixed effects analysis (with restricted maximum likelihood) of the relationship between x and y-coordinates (respectively) and ranks of turning (fixed effect). As random effect, we have intercept for the experiment as well as by-experiment random slopes to account for the non-independence of the U-turns within a group size. The examinations of residuals did not reveal any obvious deviations from homoscedasticity or normality. P-values were obtained by likelihood ratio tests of the full model with the fixed effect against the null model with intercept and random effect only. The slope estimated with restricted maximum likelihood and the result of the likelihood ratio tests are reported in Tables S3 and S2.

Table S2: Results of the linear mixed models fitted on each group size to test the effect of the rank of turning on the position regarding the y-coordinates at initiation of collective U-turns (see Figure S7). Collective U-turns with missing positions at initiation have been discarded.

Group size	Number of collective U-turns considered	Estimated slope (±se)	χ^2	p-value
2	1114	-34.83 ±4.89	18.72	< 0.001
4	1655	-25.09 ±1.78	29.19	< 0.001
5	472	-17.42 ±2.59	18.03	< 0.001
8	272	-11.34 ±0.76	45.25	< 0.001
10	33	-11.52 ±2.25	11.85	< 0.001

Table S3: Results of the linear mixed models fitted on each group size to test the effect of the rank of turning on the position regarding the x-coordinates at initiation of collective U-turns (see Figure S7). Collective U-turns with missing positions at initiation have been discarded.

Group size	Number of collective U-turns considered	Estimated slope (±se)	χ^2	p-value
2	1114	-12.04 ±4.89	5.13	0.02
4	1655	-4.04 ±0.41	56.97	< 0.001
5	472	-2.04 ±1.28	2.27	0.13
8	272	-0.95 ±0.42	19.44	< 0.001
10	33	-0.19 ±0.50	0.14	0.71

1.6 Model optimisation

Two models are discussed in this article. The *homogeneous model* with one parameter J for all group sizes and the *heterogeneous model* with one parameter J_n for each group size n. Simulations of the homogeneous model have been used for the figures presented in the main text. Each model has been simulated with two topologies, the reference topology (Figure S10, used for the figures presented in the main text) and the alternative topology (Figure S12). This section details the optimisation procedure of both models.

Table S4: Results of the linear mixed models fitted on each group size to test the effect of the rank of turning on the position regarding the y-coordinates when the individual turns (see Figure S8). Collective U-turns with missing positions at fish turns have been discarded.

Group size	Number of collective U-turns considered	Estimated slope (±se)	χ^2	p-value
2	1114	-5.86 ± 2.43	4.87	0.03
4	1655	-5.54 ± 1.51	8.85	< 0.001
5	472	-2.13 ± 1.71	1.29	0.26
8	272	1.40 ± 2.47	0.27	0.60
10	33	1.32 ± 1.59	0.54	0.463

Table S5: Results of the linear mixed models fitted on each group size to test the effect of the rank of turning on the position regarding the x-coordinates when the individual turns (see Figure S8). Collective U-turns with missing positions at fish turns have been discarded.

Group size	Number of collective U-turns considered	Estimated slope (±se)	χ^2	p-value
2	1114	-9.68 ± 4.00	5.00	0.03
4	1655	-3.91 ± 0.28	36.24	< 0.001
5	472	-3.51 ± 2.19	2.42	0.12
8	272	-2.98 ± 0.75	9.22	< 0.001
10	33	-5.36 ± 1.26	16.78	< 0.001

1.6.1 Homogeneous model

For given J and ϵ, we compute numerically the prediction for the number of collective U-turns u'_n for a group of size n made during T' Monte-Carlo time steps. We define the error function

$$\Delta = \sum_n (t_{f,n} - t_0 t'_{f,n})^2 \tag{S3}$$

with $t_{f,n}$ the empirical average time required for the last fish to turn during collective U-turns and $t'_{f,n}$ the average time required for the last agent to turn during collective U-turns in simulations. t_0 has the dimension of a time and translates Monte-Carlo time (i.e. number of Glauber updates) into actual experimental minutes, and is determined by minimising the error Δ, i.e. by solving the equation $\frac{\partial \Delta}{\partial t_0} = 0$.

1.6.2 Heterogeneous model

We developed an alternative model in which the coupling constant J_n changes with the group size while ϵ is kept fixed.

For each group size n, we find $\hat{t}_{0,n}$ that minimises $\Delta = (t_{f,n} - \hat{t}_{0,n} t'_{f,n})^2$. We find J_n by minimising:

$$\Delta = [\log(\tau_n) - \log(\hat{t}_{0,n}\tau'_n)]^2, \tag{S4}$$

with $\tau_n = \frac{u_n}{T_n}$ the empirical rate of collective U-turns (with T_n the total duration of all the experiments of the group size n, in minutes) and $\tau'_n = \frac{u'_n}{T'}$ the rate of collective U-turns in simulations. Doing so leads to a perfect agreement between model predictions and empirical data regarding the frequency of the collective U-turns. We find $t_{0,n} = \hat{t}_{0,n} t_{c,n}$, with $t_{c,n}$ a scaling parameter that minimises $\Delta = (\hat{t}_{0,n} t'_{f,n} - \hat{t}_{0,r} t'_{f,r})^2$, with $\hat{t}_{0,r} t'_{f,r}$ the average time (in seconds) for the n-th fish to turn in the group size of reference (here $r = 10$ fish). We eventually set $\tau'_n = \frac{u'_n}{T' t_{0,n}}$.

1.7 Model implementation

The models have been implemented in R (and run with R 3.3.1) with a C++ subroutine using the package `Rcpp` [4, 5].

2 Supplemental figures

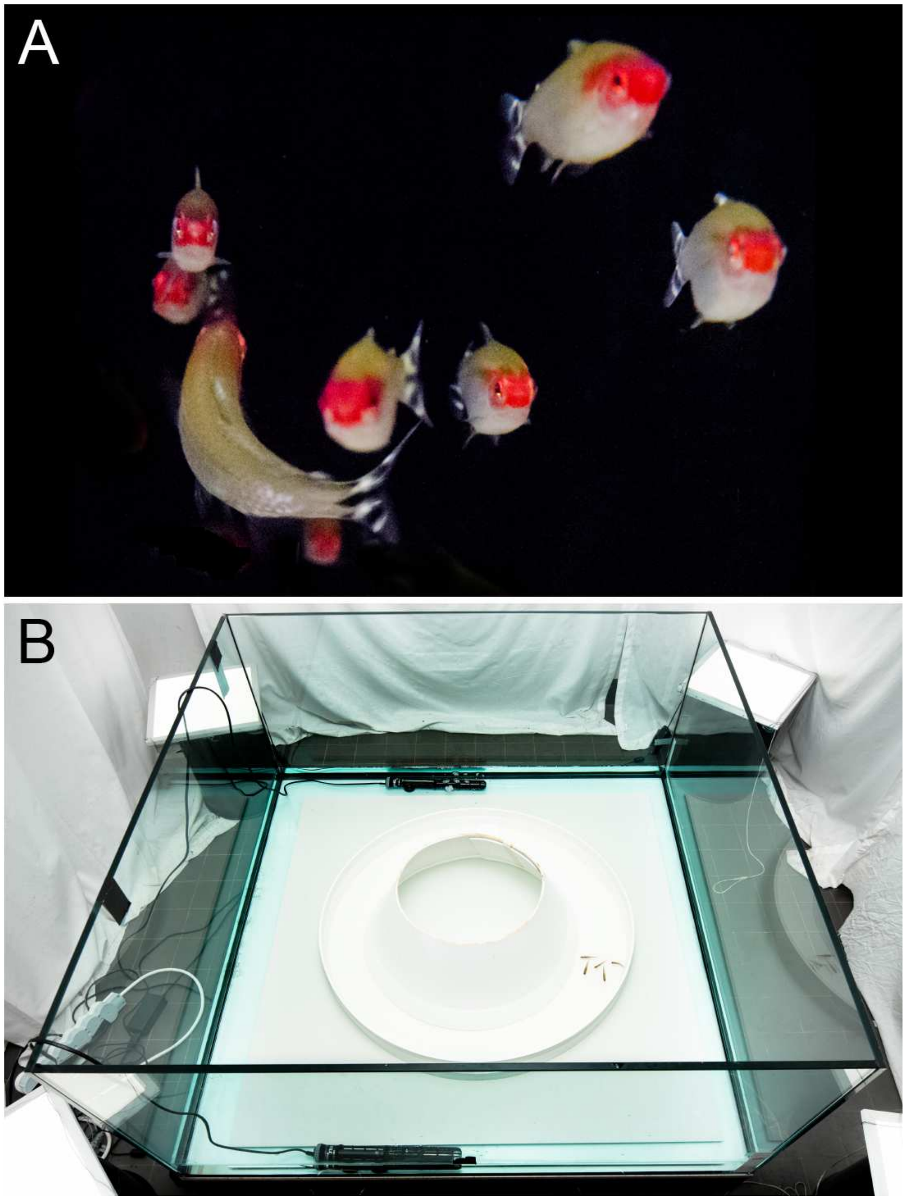

Figure S1: Experimental set-up. (A) A photo of a spontaneous U-turn initiated by a single fish in a group of eight *Hemigrammus rhodostomus* fish, (B) Experimental ring-shaped tank, credits to David Villa ScienceImage/CBI/CNRS, Toulouse.

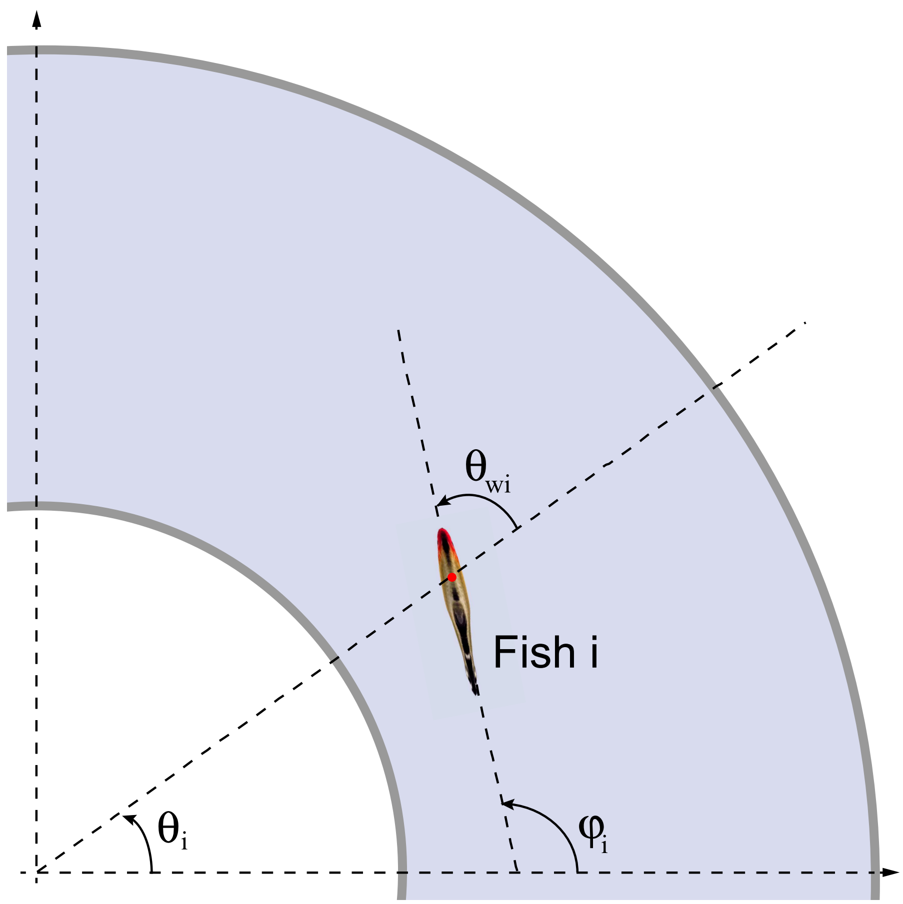

Figure S2: Variables used to describe the position, heading and relative orientation of fish relative to the wall in the experimental set-up: θ_i is the angle formed by the position vector of fish i and the horizontal line, φ_i is the heading of fish i, and θ_{wi} is the angle of incidence of fish i relative to the wall w.

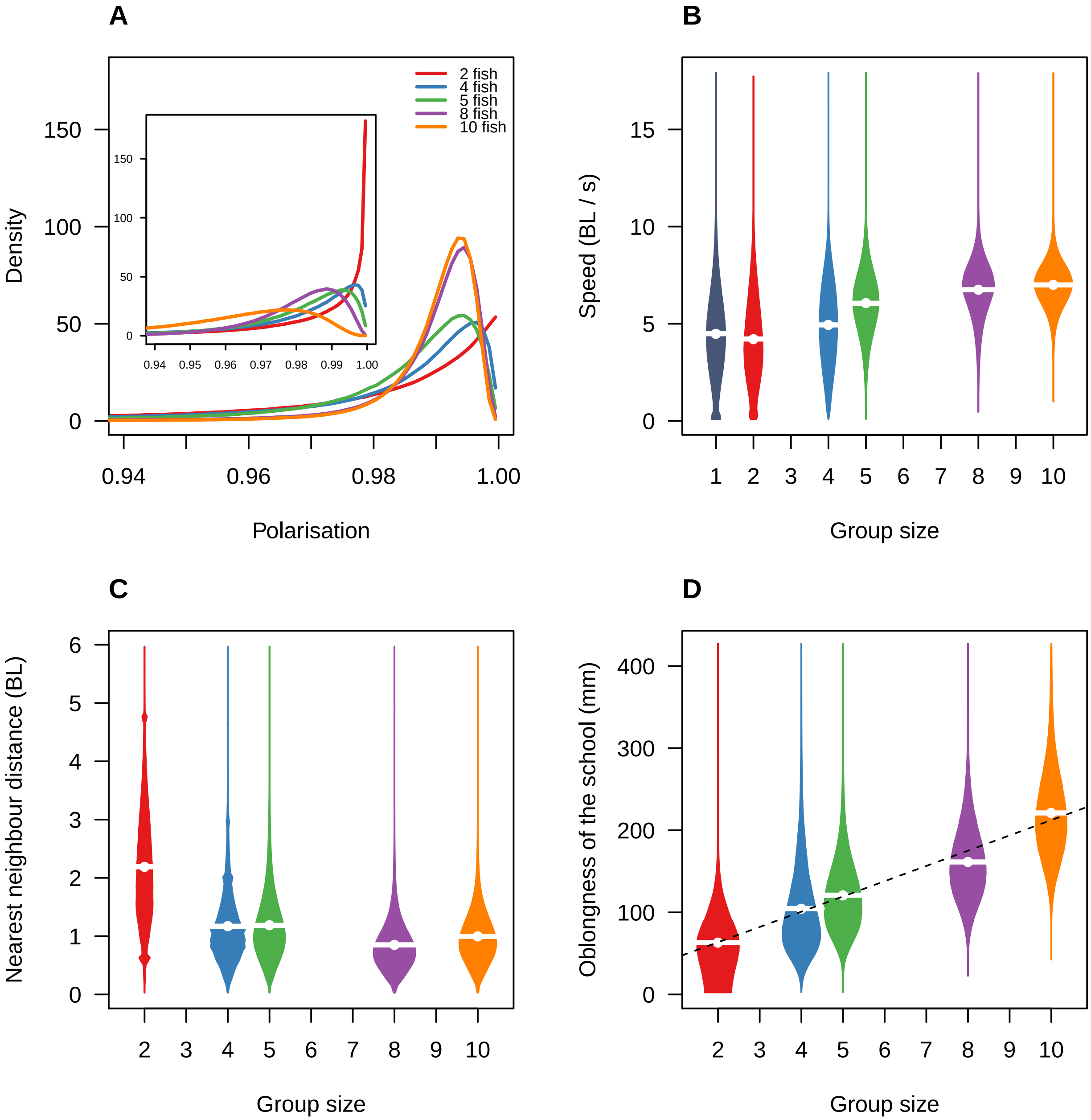

Figure S3: Influence of group size on internal structure, speed and shape of the schools. A) Distribution of polarisation, measured as the absolute value of the degree of alignment with the wall a_i and (inset) as the order parameter $\Phi = \sqrt{(\sum_N (\cos\phi_i)^2 + \sum_N (\sin\phi_i)^2)/N}$. Both parameters tends to 0 when the group is disordered and to 1 when the group is perfectly ordered. B) Distribution of the speed of the group, averaged over the speed of each individual, at each time, as a function of group size. C) Distribution of the nearest neighbour distance, measured on each individual, at each time, as a function of group size. D) Distribution of the oblongness of the group, measured on each frame as the maximum distance between positions of fish projected on the axis tangent to the swimming direction of the centre of mass of the group, as a function of group size. Dashed line stands for fitted linear model, $R^2 = 0.98$. B, C and D are violin plots, showing the rotated and mirrored histograms of the respective random variable. White belts stand for the mean.

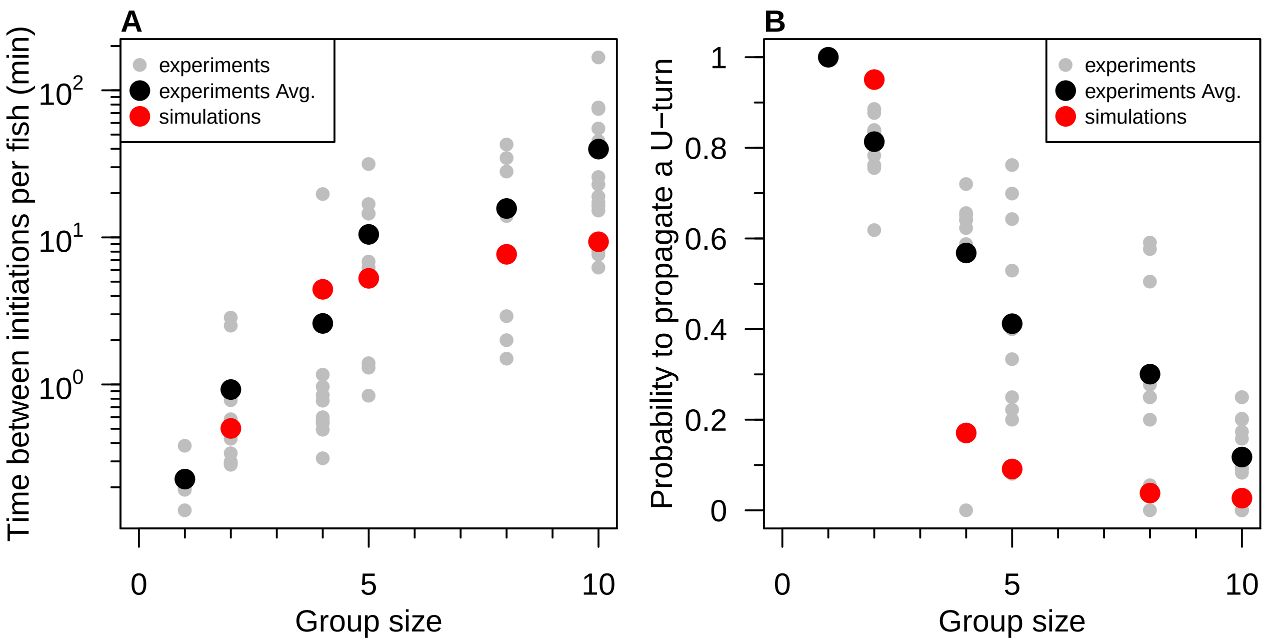

Figure S4: A) Time between U-turn initiation (failed or fully propagated) per fish as a function of group size. B). Probability that an initiated U-turn is fully propagated as a function of group size (see equations S1 and S2).

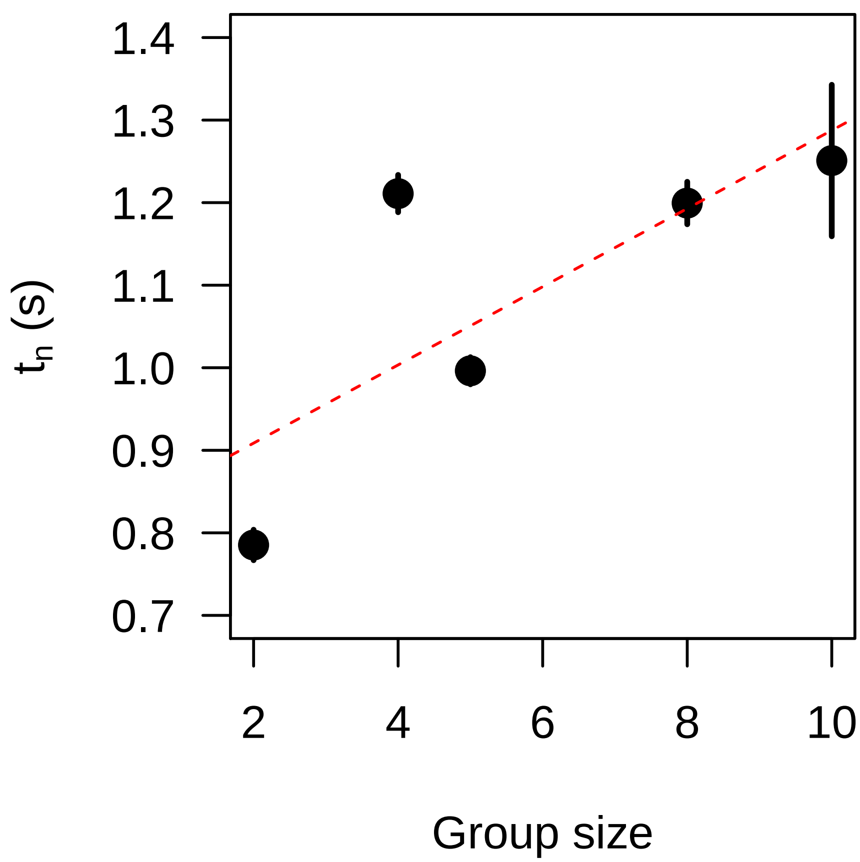

Figure S5: Mean ± sd of the bootstrapped sample of the scaling coefficient t_n used in Figures 1C and 1D; red dashed line stands for a fitted linear model, $R^2 = 0.59$.

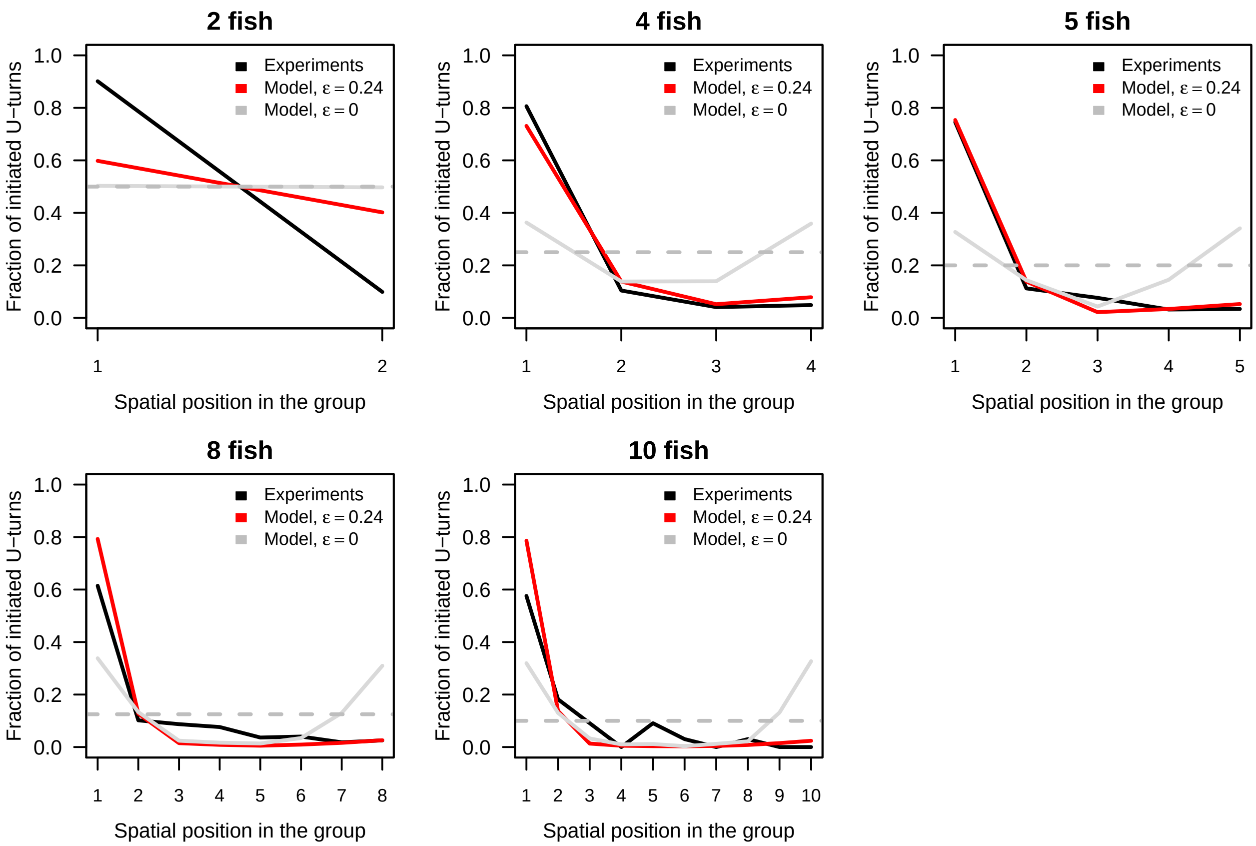

Figure S6: Spatial position of the U-turn initiator in groups of 2, 4, 5, 8 and 10 fish, in empirical data (black lines) and simulations (red and solid grey lines for $\epsilon = 0.24$ and $\epsilon = 0$ respectively). Dashed grey lines show the probability of initiating a collective U-turn without any effect of the spatial position (i.e. $1/n$). For $\epsilon \neq 0$ (red lines), simulations reproduce the inhibition of initiation of collective U-turns at the rear of the groups. Without perception anisotropy and asymmetry ($\epsilon = 0$, solid grey lines), U-turns are initiated by the fish that have fewer influential neighbours (in our simulations, those are the fish at the boundary of the group – all individuals would have the same probability to initiate a U-turn with periodic boundary conditions) and propagated to their neighbours without favouring any direction.

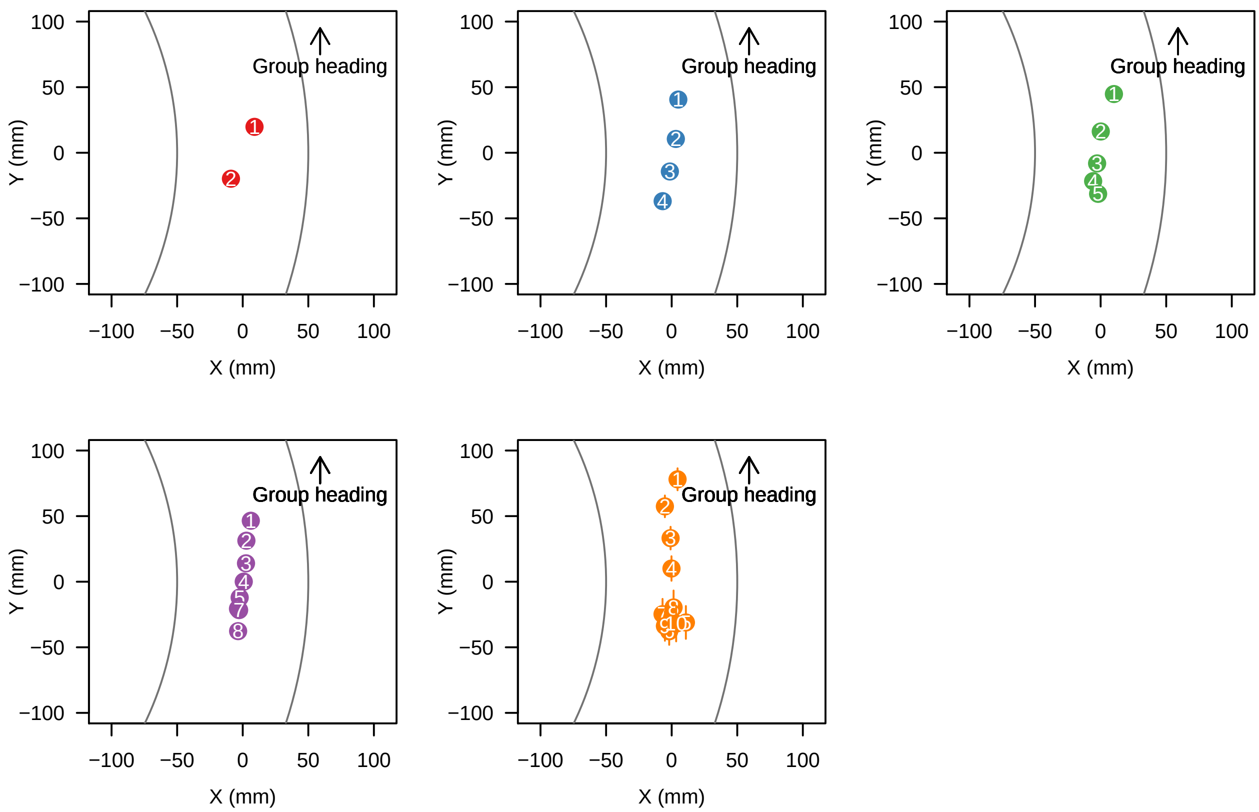

Figure S7: Average positions at U-turn initiation of individual that turn subsequently, indicated by their ranks of turning (where rank 1 is the initiator of the U-turns) in experiments for groups of 2, 4, 5, 8 and 10 fish. Positions have been corrected so that all groups move in the same direction, with the outer wall at their right hand-side. Error bars indicate the standard error of the x and y-coordinates (smaller than the circles if not visible). The origin of the coordinate system is set to the centroid of the average positions of individuals. Statistical tests regarding the effect of the ranks of turning on the x and y-positions are reported in Tables S3 and S2.

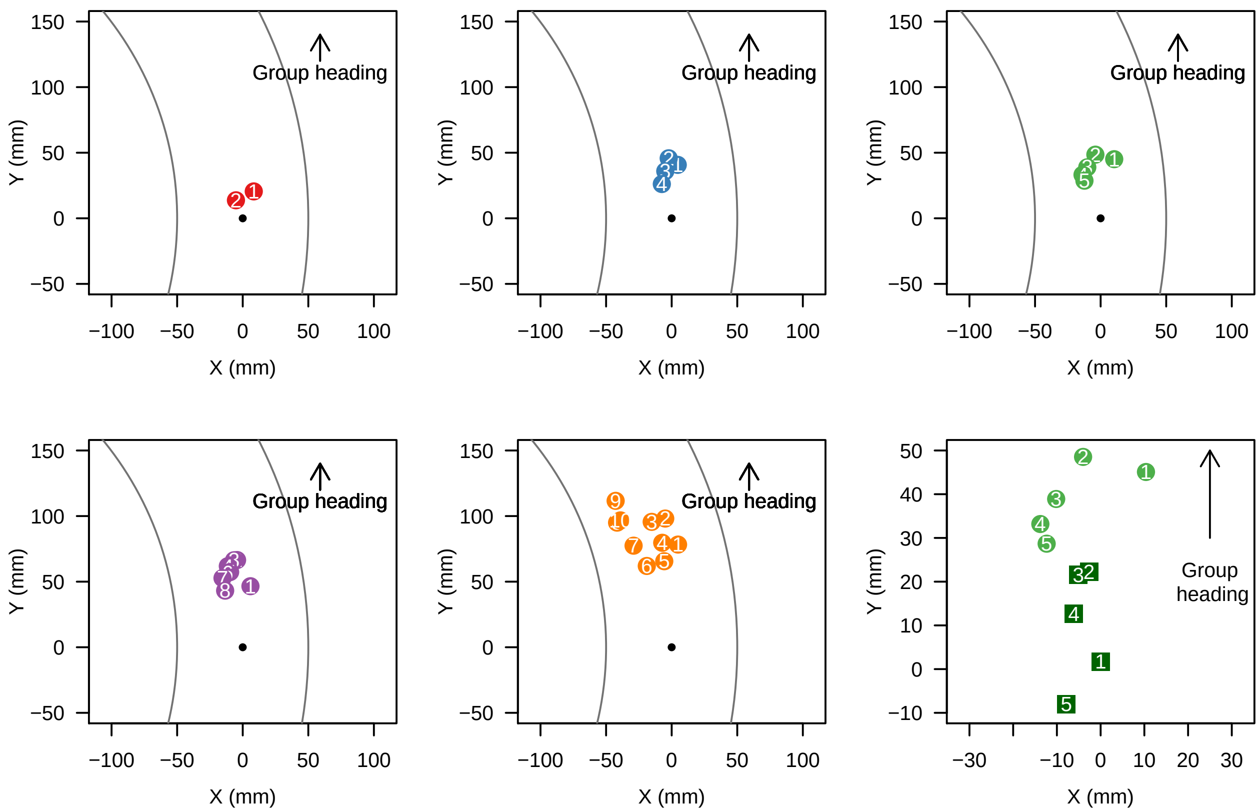

Figure S8: Average positions of individual that turn subsequently, indicated by their ranks of turning (where rank 1 is the initiator of the U-turns) in experiments for groups of 2, 4, 5, 8 and 10 fish. Positions have been corrected so that all groups move in the same direction, with the outer wall at their right hand-side. Error bars indicate the standard error of the x and y-coordinates (smaller than the circles if not visible). The origin of the coordinate system (black dot) is set to the centroid of the average positions of individuals at the initiation of the collective U-turns. Note that the larger the group, the more the fish travel since the start of the U-turn, because fish swim faster (Figure S3B) and a U-turn takes a longer time in larger groups (Figure S5). Statistical tests regarding the effect of the ranks of turning on the x and y-positions are reported in Tables S5 and S4. Bottom-right panel is a zoom of the top-right panel (i.e. for 5 fish) where dark-green squares are the centroid of the positions of all fish for each turning rank.

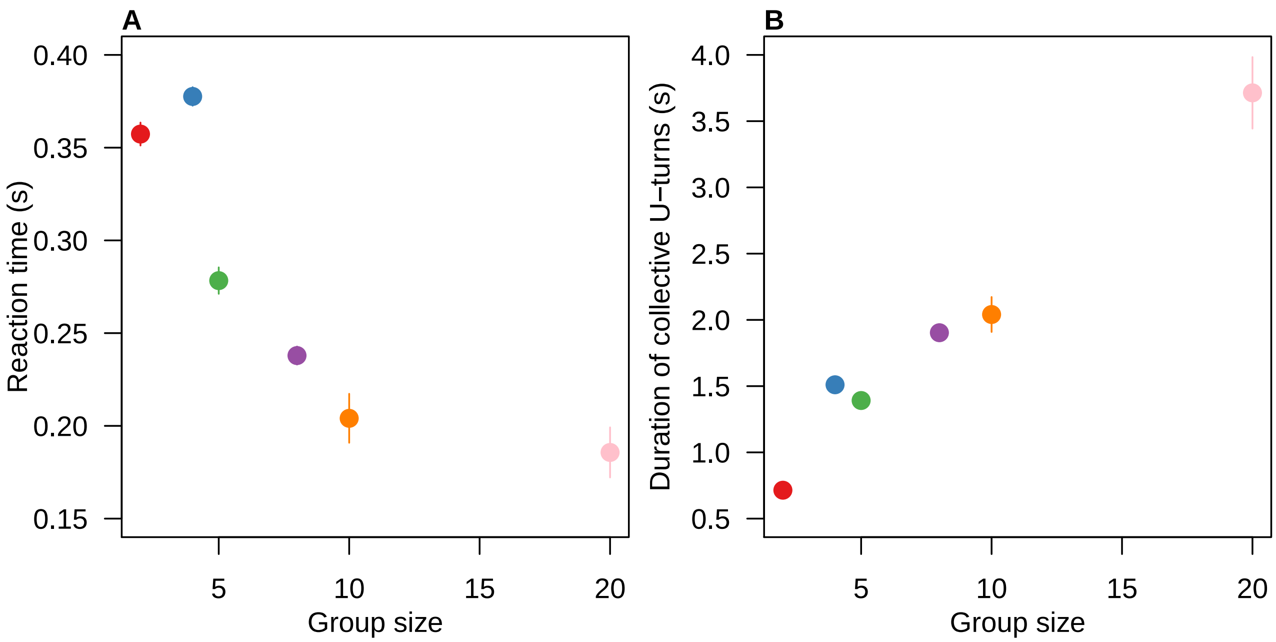

Figure S9: A). Reaction time measured as the average time interval between subsequent individuals making a U-turn (±s.e.) as a function of group size. The average time interval between subsequent individuals making a U-turn is computed as the average duration of collective U-turns divided by the group size. B). Average duration of the collective U-turns (±s.e.) as a function of group size. See Figure S19 for the distributions.

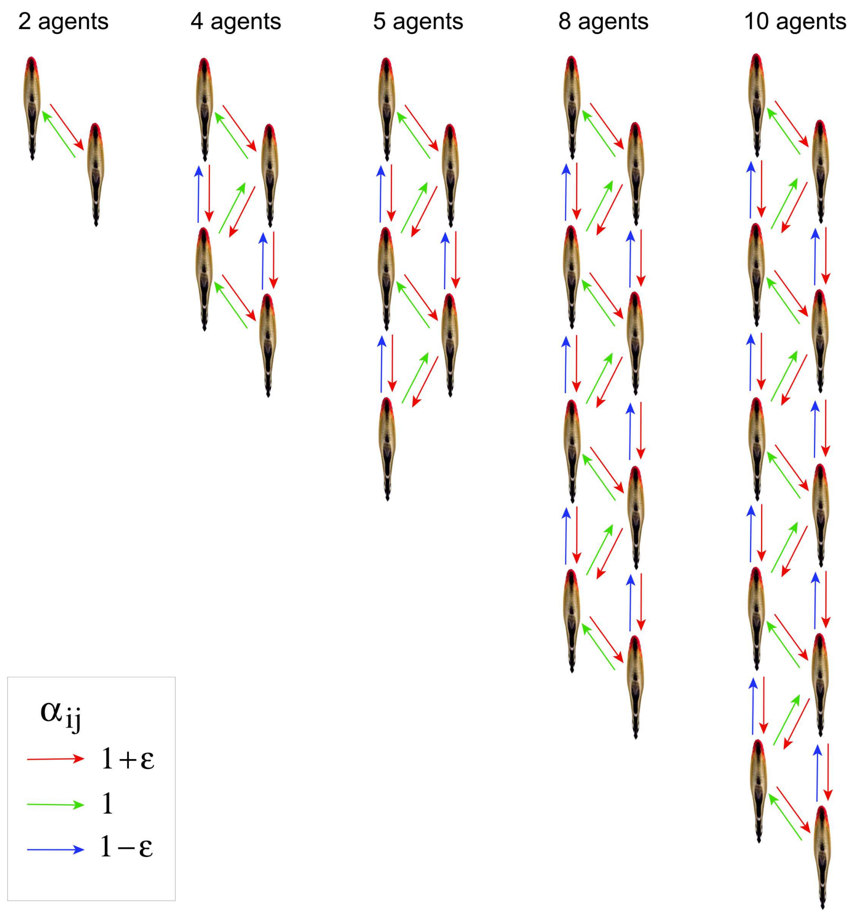

Figure S10: Topology of the interaction network of reference in the simulations for different group sizes. Arrows indicate interactions going from the influencing agent to the influenced one. The colour of the arrow refers to the weight of the interaction, namely $\alpha_{ij} = 1 + \epsilon$ (red arrow), $\alpha_{ij} = 1$ (green arrow) and $\alpha_{ij} = 1 - \epsilon$ (blue arrow). The number of influencing neighbours of a focal agent can be derived from the number of pairs of arrows connected to the agent. For instance, in groups of 5 agents, each agent has, respectively (from front to back) 2, 3, 4, 3 and 2 influencing neighbours.

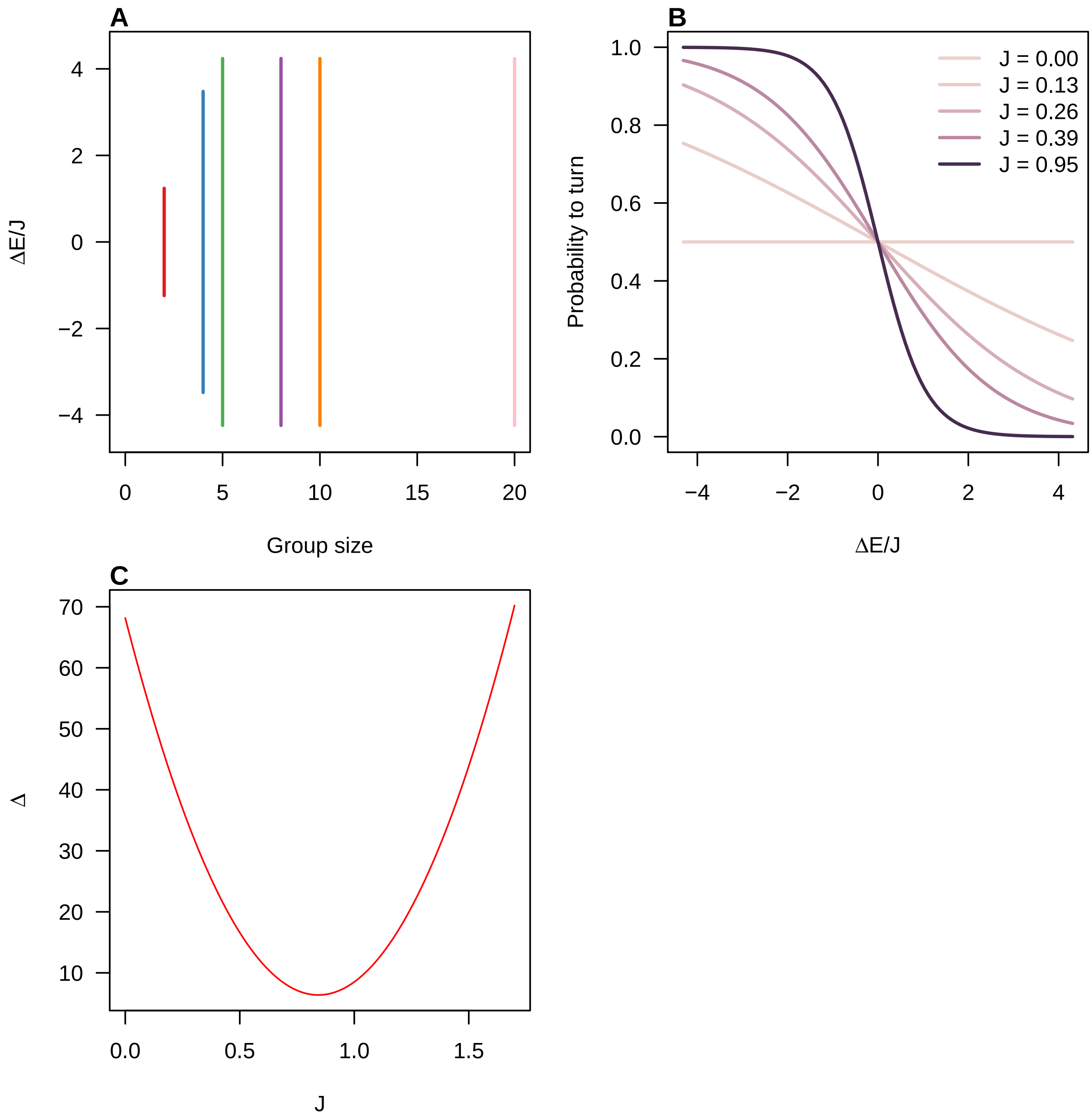

Figure S11: Influence of nonlinearity in the model, with the anisotropy parameter $\alpha = 0.24$. A) Range of $\Delta E/J$ values given $\alpha = 0.24$ for each group size. B) Probability P to accept an individual U-turn (see main text, equation 3.5) for different values of J (including $J = 0.95$, used in the fitted model). This figure shows that on the considered range $\Delta E/J$, the acceptance probability chosen in the main text ($J = 0.95$) is strongly non-linear. C) Error $\Delta = \sum_n [\log(\tau_n) - \log(t_0 \tau'_n)]^2$ as a function of the coupling constant J. Values of J giving a non-linear acceptance probability (B) are necessary to improve the agreement between the average time between collective U-turns in experiments and simulations.

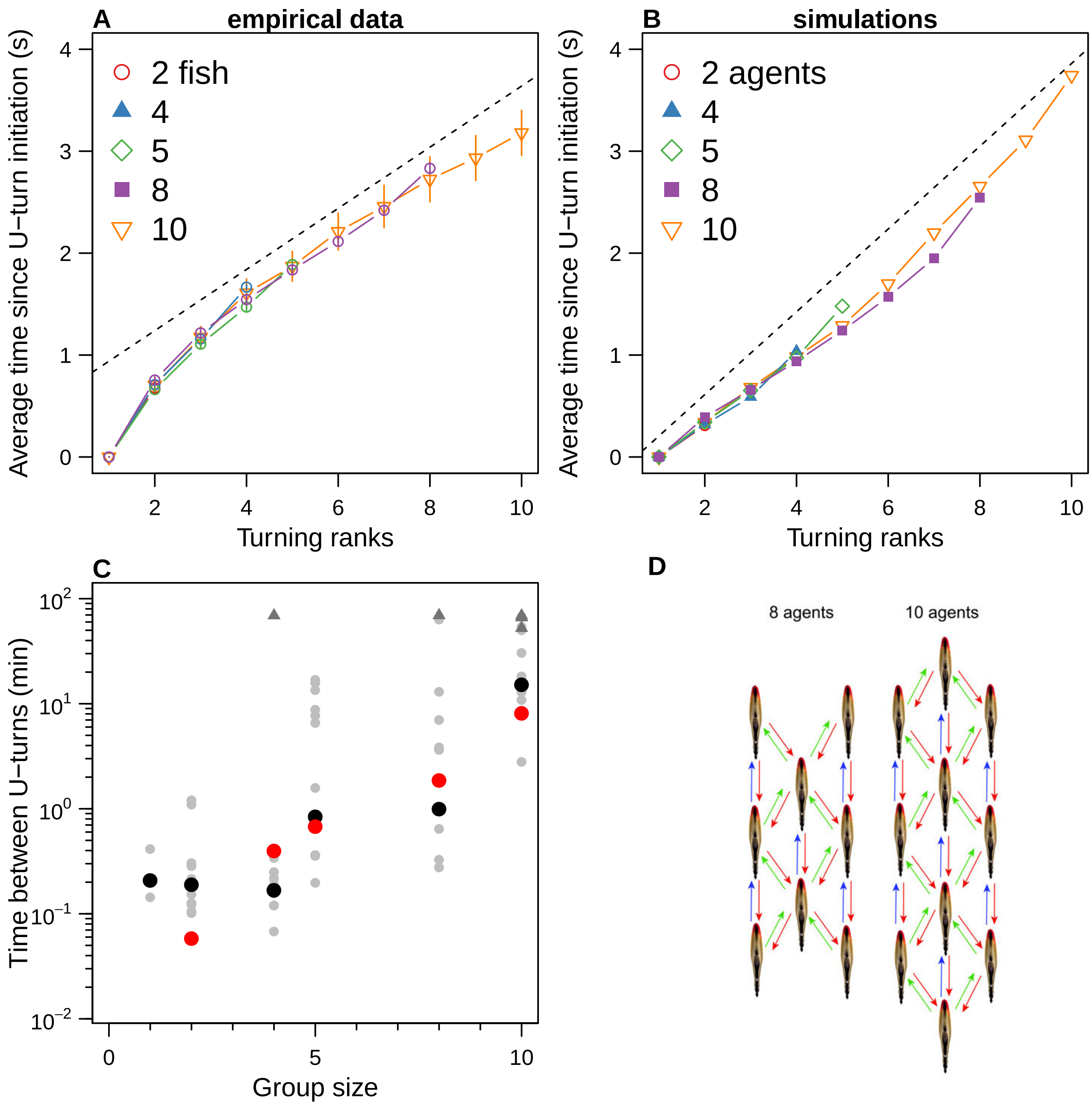

Figure S12: Results of simulations of the model with constant J (for all group size n) and an alternative topology for the groups of size $n = 8$ and 10, with $J = 0.69$. A) Average time interval since the beginning of a collective U-turn as a function of turning rank and group size in empirical results. B) Average time interval since the beginning of a collective U-turn as a function of turning rank and group size in simulations with the alternative topology. C) Average time between two consecutive collective U-turns as a function of group size in experiments and simulations. Grey dots stand for the average time between collective U-turns $\rho_n^l = \frac{T_n^l}{u_n^l}$ in each experiment, grey triangles experiments without collective U-turns, black dots the average $\rho_n = \frac{T_n}{u_n}$ over all experiments and red dots the average $\rho_n' = \frac{T' t_0}{u_n'}$ in simulations. D). Topologies of the interaction network for 8 and 10 fish in the alternative topology.

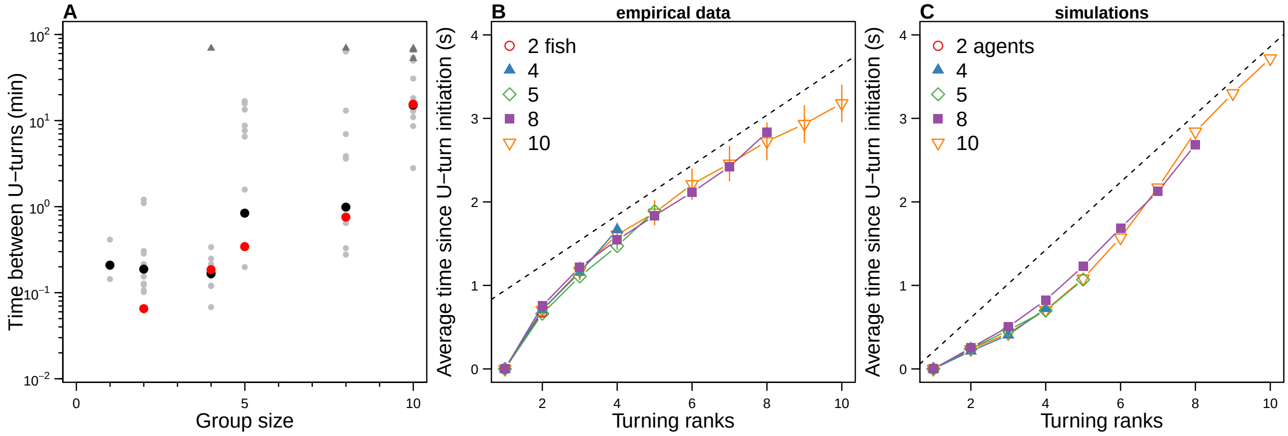

Figure S13: Results of simulations of the model with n-dependent J_n and the reference topology of Figure S10. A) Average time between two consecutive collective U-turns as a function of group size in experiments and simulations. Grey dots stand for the average time between collective U-turns $\rho_n^l = \frac{T_n^l}{u_n^l}$ in each experiment, grey triangles experiments without collective U-turns, black dots the average $\rho_n = \frac{T_n}{u_n}$ over all experiments and red dots the average $\rho'_n = \frac{T' t_{0,n}}{u'_n}$ in simulations. B) Average time interval since the beginning of a collective U-turn as a function of turning rank and group size in empirical results. C) Average time interval since the beginning of a collective U-turn as a function of turning rank and group size in simulations.

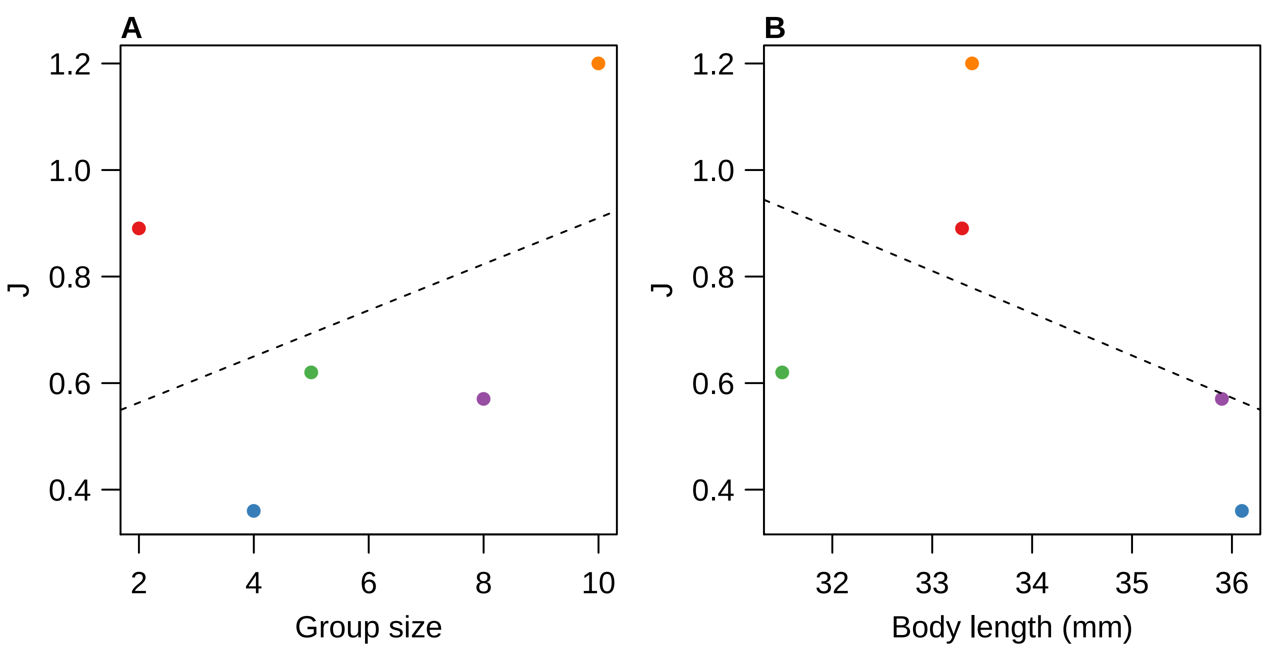

Figure S14: Parameters of simulations of the model with n-dependent J_n and the reference topology. A) Fitted values of J_n as a function of the group size. Dashed line shows fitted linear regression ($R^2 = 0.18$, $t = 0.82$, p-value $= 0.47$). B) Fitted values of J_n as a function of the average body length of each group size. Dashed line shows fitted linear regression ($R^2 = 0.23$, $t = -0.94$, p-value $= 0.42$).

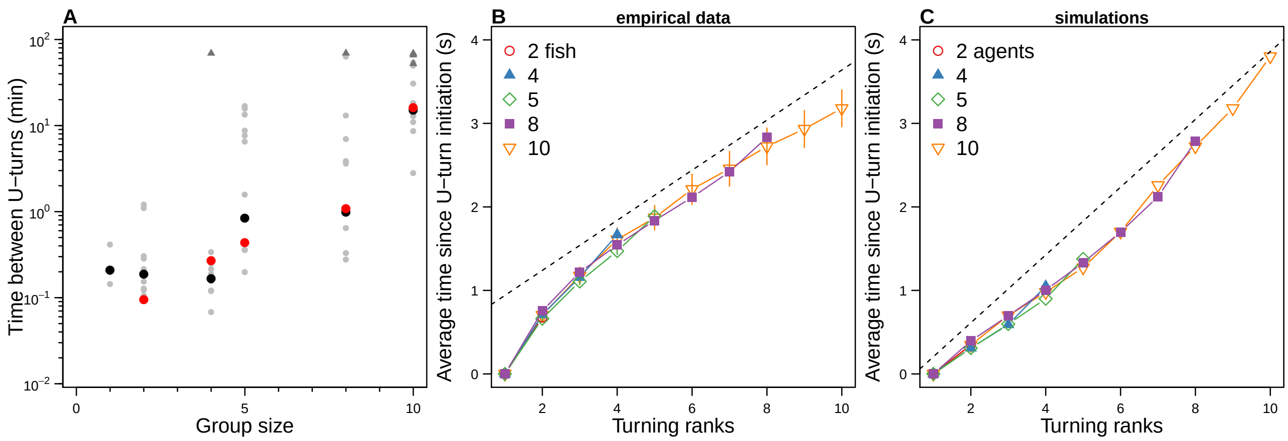

Figure S15: Results of simulations of the model with n-dependent J_n and the alternative topology of Figure S12. A) Average time between two consecutive collective U-turns as a function of group size in experiments and simulations. Grey dots stand for the average time between collective U-turns $\rho_n^l = \frac{T_n^l}{u_n^l}$ in each experiment, grey triangles experiments without collective U-turns, black dots the average $\rho_n = \frac{T_n}{u_n}$ over all experiments and red dots the average $\rho'_n = \frac{T' t_{0,n}}{u'_n}$ in simulations. B) Average time interval since the beginning of a collective U-turn as a function of turning rank and group size in empirical results. C) Average time interval since the beginning of a collective U-turn as a function of turning rank and group size in simulations.

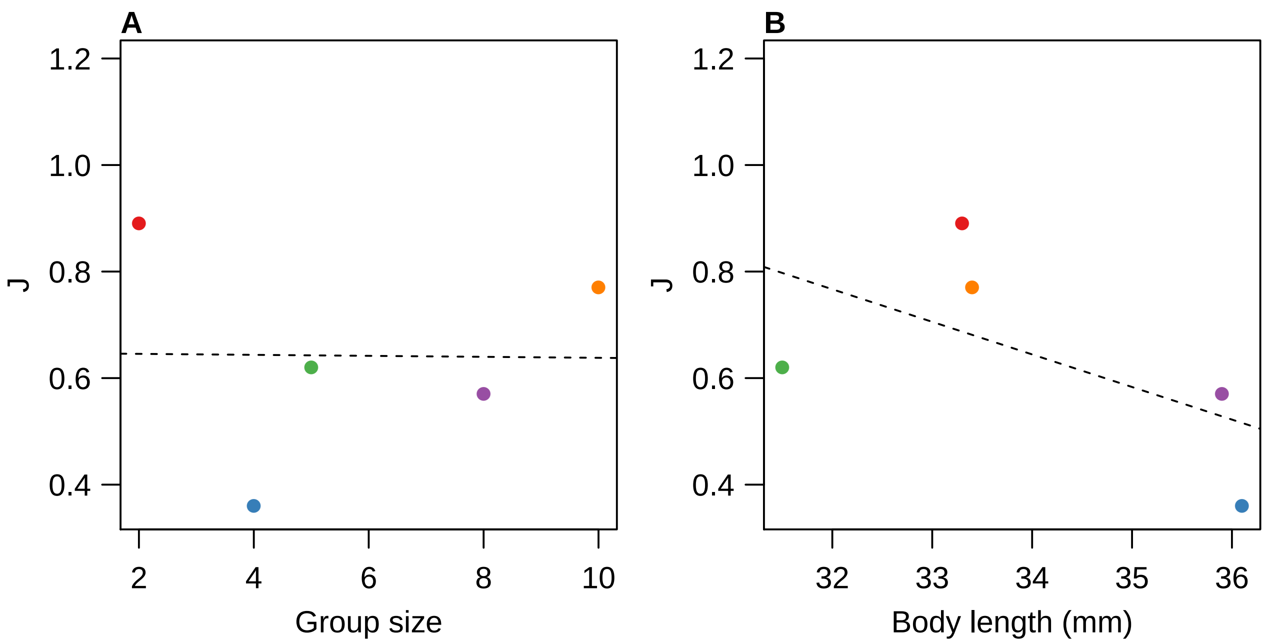

Figure S16: Parameters of simulations of the model with n-dependent J_n and the alternative topology. A) Fitted values of J_n as a function of the group size. Dashed line shows fitted linear regression ($R^2 \approx 0$, $t = -0.03$, p-value $= 0.98$). B) Fitted values of J_n as a function of the average body length of each group size. Dashed line shows fitted linear regression ($R^2 = 0.35$, $t = -1.26$, p-value $= 0.30$).

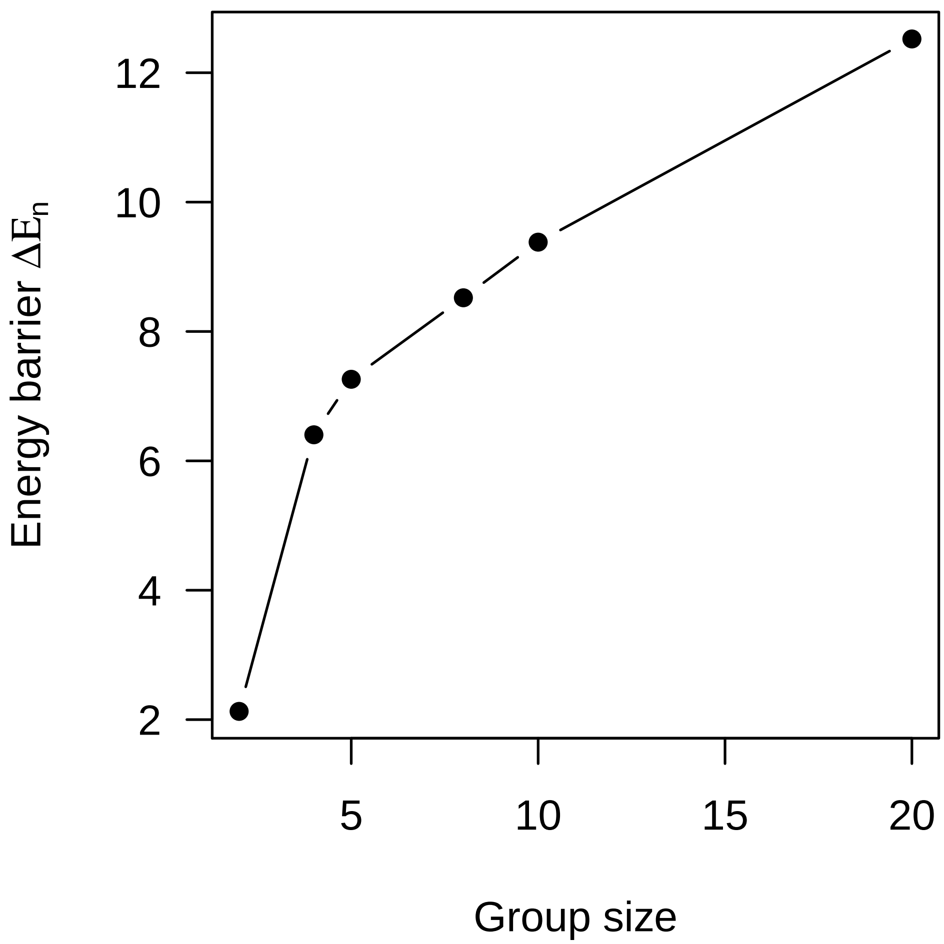

Figure S17: Energy barrier as a function of group size. The energy barrier ΔE_n for a group of size n is calculated in simulations as the difference between the average maximum of energy reached during U-turns and the reference energy when all agents are heading in the same direction.

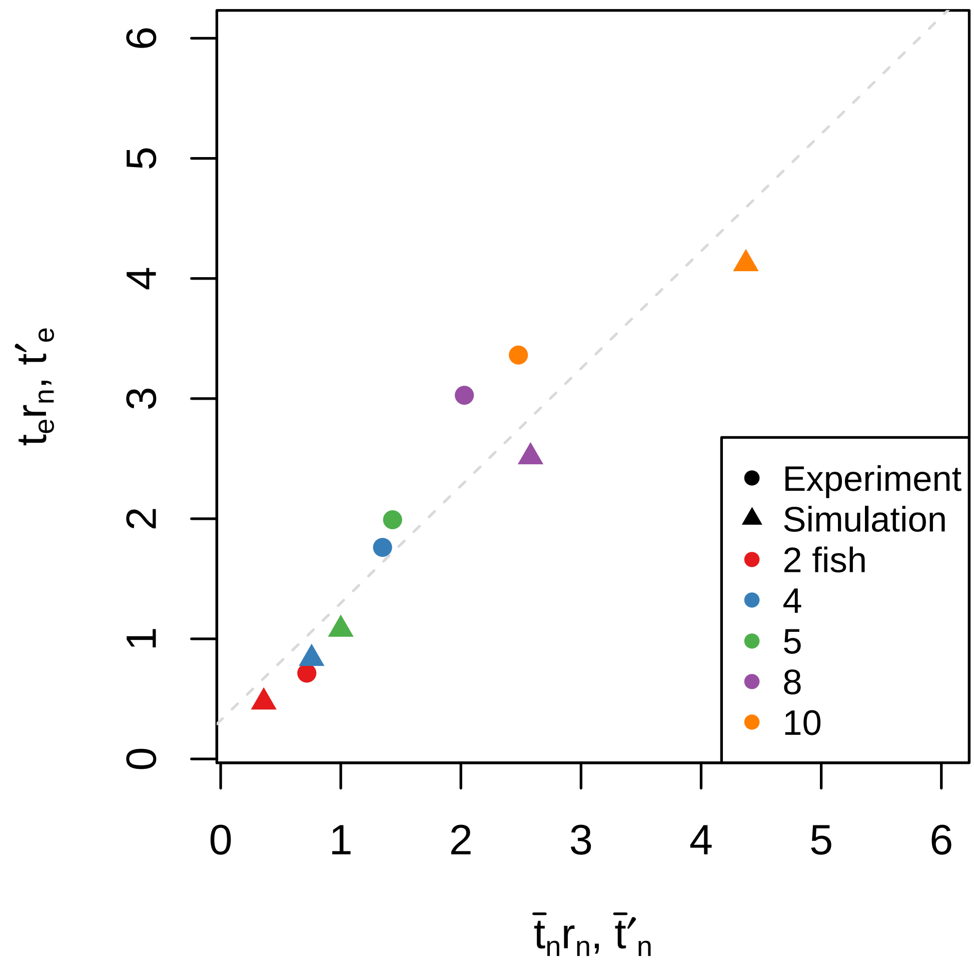

Figure S18: Duration of the collective U-turns, both in experiments $t_e r_n$ and the model t'_e as a function of the scaling coefficients ($\bar{t}_n r_n$ in experiments and $\bar{t}'_n$ in the model) obtained from the data scaling shown on Figure 4 of the main text (see section 1.4).

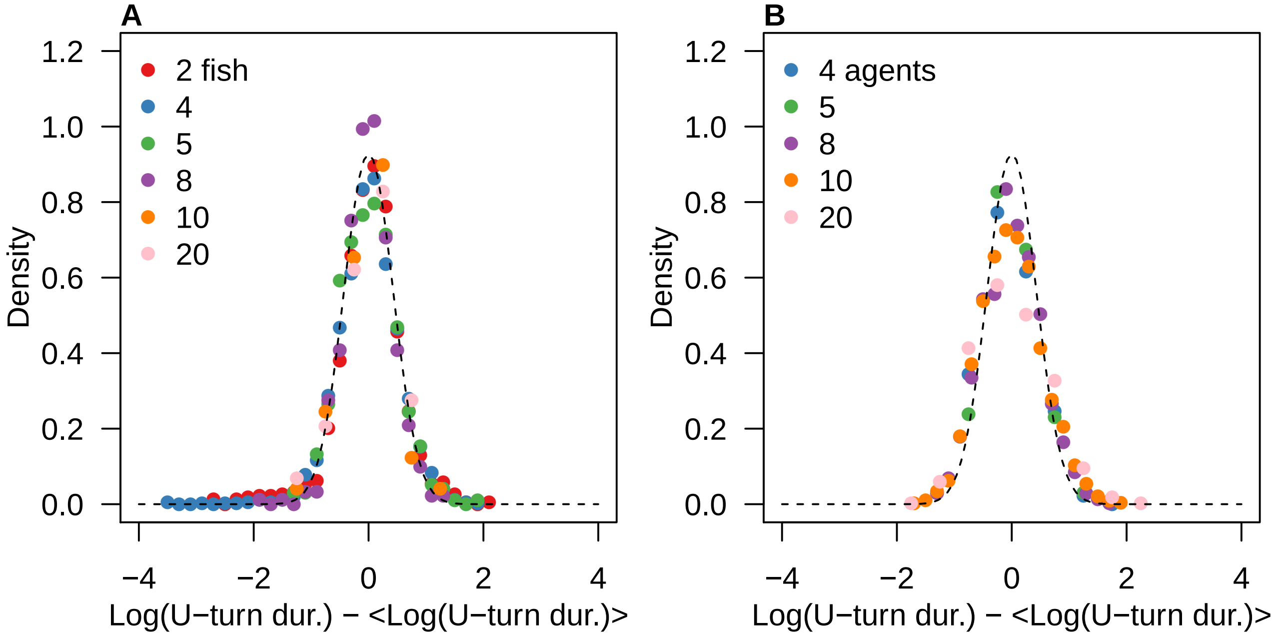

Figure S19: Probability distribution of the durations of U-turns normalised by the average duration, for each group size in experiments (A) and numerical simulations of the model (B). Dashed line is the probability density function of the Normal distribution $\mathcal{N}(\mu = 0, \sigma = 0.43)$.

Movie S1: This thumbnail shows a collective U-turn in a group of 8 fish. The associated movie shows one collective U-turn per group size, namely 1 fish and groups of 2, 4, 5, 8 and 10.

References

[1] A. Pérez-Escudero, J. Vicente-Page, R. C. Hinz, S. Arganda, and G. G. de Polavieja, "id-Tracker: Tracking individuals in a group by automatic identification of unmarked animals," *Nat. Methods*, pp. 743–748, 2014.

[2] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2016. [Online]. Available: https://www.R-project.org/

[3] D. Bates, M. Mächler, B. Bolker, and S. Walker, "Fitting linear mixed-effects models using lme4," *Journal of Statistical Software*, vol. 67, no. 1, pp. 1–48, 2015.

[4] D. Eddelbuettel and R. Francois, "Rcpp: Seamless r and c++ integration," *Journal of Statistical Software, Articles*, vol. 40, no. 8, pp. 1–18, 2011. [Online]. Available: https://www.jstatsoft.org/v040/i08

[5] D. Eddelbuettel, *Seamless R and C++ integration with Rcpp*. Springer, 2013.