Aggregation rule in animal collectives dynamically changes between majority and minority influence

Francisco J.H. Heras,[1*] Francisco Romero-Ferrero,[1]
Robert C. Hinz,[1] Gonzalo G. de Polavieja,[1*]

[1] Champalimaud Research, Champalimaud Centre for the Unknown, Lisbon, Portugal
[*] Correspondence: francisco.heras@neuro.fchampalimaud.org
/ gonzalo.depolavieja@neuro.fchampalimaud.org

Abstract

Majority-voting or averaging the estimations made by the individuals in a collective are simple rules that can effectively aggregate knowledge under some ideal conditions. However, these rules can catastrophically fail in the frequent situation in which a minority brings knowledge to a collective. Aggregation rules should ideally use majorities or averages when most or all members have similar information and focus on a minority when it brings new relevant information to the group. Here we turned to fish schools to test whether aggregation rules that have evolved over hundreds of millions of years can use this flexible aggregation. We tracked each animal in large groups of 60, 80 and 100 zebrafish, *Danio rerio*, with a newly developed method. We used the trajectories to train deep attention networks and obtained a model that is both predictive and insightful about the structure of fish interactions. A six-dimensional function describes the focal-neighbour interaction and a four-dimensional function how information is aggregated. The aggregation function shows that each animal sometimes averages approximately 25 neighbours and sometimes focuses on fewer animals down to effectively a single one, and that it can rapidly shift between these extremes depending on the relative positions and velocities of local neighbours. Animal collectives could thus avoid the limitations of simple rules and instead flexibly shift from average many to follow few or one individual.

1 Introduction

Aggregation rules that weigh all members in a collective equally can give estimations that are better on average than choosing one member at random [1, 2]. However, this is a mathematical certainty only under some particular conditions [3, 4]; e.g. when all individuals in a collective are noisy estimators but better than random and statistically independent [1, 5]. There is an intense debate on whether the conditions under which these simple rules are guaranteed to work are useful to defend a truth-tracking version of majority or plurality voting in society [6, 7, 4, 8]. These rules might still work when knowledge is in a minority, for example when an ignorant majority chooses with equal probability among all available options and the informed minority biases towards the correct answer [5, 9, 10]. However, even if this is the case, the effect of a minority is small and thus very sensitive to noise.

Many models of interactions in collective animal behavior also use an average-across-neighbours approach as a simplifying assumption [11, 12, 9, 13, 14, 15, 16] or other operations that also aggregate all individuals equally [17, 18, 19, 20, 21]. However, there are models, like the many-eyes model for predator detection, explaining that, after some individuals detect a predator, the rest do not aggregate all members equally but focus instead on the informed minority [22, 23, 24, 25, 26].

An ideal aggregation system would have these two cases —averaging or following very few individuals— as extremes, shifting between them to match the changing knowledge distribution in the group [4]. We set out to test for this type of aggregation rule using high-quality tracking data of large animal groups of zebrafish, *Dario rerio*, obtained using the recently developed idtracker.ai [27], and modelling techniques using deep attention networks [28, 29, 30].

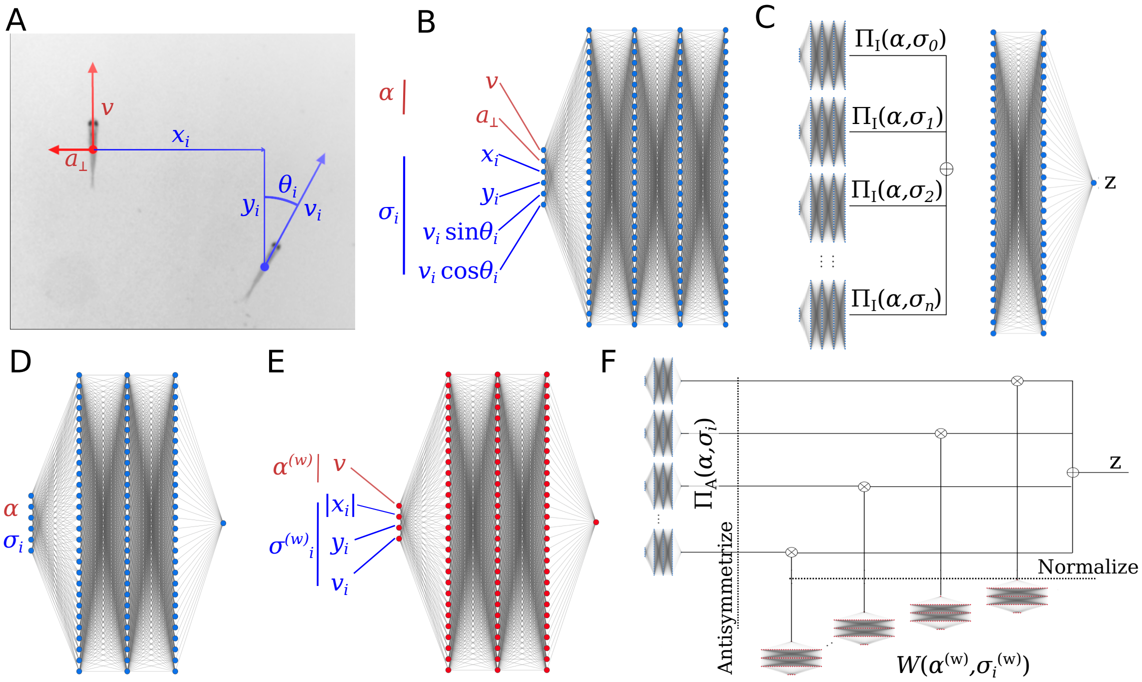

Figure 1: **Deep-learning a model of collective behaviour**. (**A**) Variables used to predict future turns. Asocial variables, those only involving the focal, in red. Social variables, those involving both the focal and a neighbour, in blue. (**B**) Pair-interaction subnetwork, receiving asocial variables α and social variables σ_i from a single neighbour i, and outputting a vector of 128 components. All pair-interaction networks share the same weights. (**C**) Interaction network, showing how the outputs of the pair-interaction subnetworks, one for each neighbour, are summed and then fed to an interaction subnetwork. The output, z is the logit of the focal fish turning right after 1 s. (**D**) Pair-interaction subnetwork of the attention network. (**E**) Aggregation subnetwork of the attention network. Same structure as D, but the input is a restricted symmetric subset of the variables and the output is passed through an exponential function to make it positive. (**F**) Attention network, showing how the inputs of the pair-interaction and attention subnetworks are integrated to produce a single logit z for the focal fish turning right after 1 s.

2 Results

2.1 Predicting the future using a deep interaction network

We recorded videos of groups of 60, 80 or 100 juvenile zebrafish, *Danio rerio*, (**Fig. 1A** for a detail of two fish and **Fig. S1** for illustrative video frames). We tracked videos using our system idtracker.ai, obtaining high-quality position, velocity and acceleration values (see **Materials and Methods**).

We used the trajectories to obtain data-driven models of fish interactions. We required our models to be predictive of the future of a focal fish in test data (video sequences not used to train the model). We found deep interaction networks, good predictors of physical systems like the motion of planets [31], to have the highest predictive power on test data. Our deep interaction network is divided in two parts: (i) n pair-interaction subnetworks, each describing the interaction of a focal fish with one of its n closest neighbors (**Fig. 1B**), and (ii) an aggregation subnetwork, aggregating the n outputs of the pair-interaction subnetworks (**Fig. 1C**, subnetwork to the right).

The inputs to the network are quantities expressed in a coordinate system centered at the focal fish and with the y-axis in the direction of the velocity of the focal (**Fig. 1A**, red). The pair-interaction subnetwork has as inputs the asocial information of the focal, α (**Fig. 1B**, red), and the social information of one neighbour i, σ_i (**Fig. 1B**, blue). The asocial information of the focal is its speed, v, tangential acceleration, $a_\parallel$, and normal acceleration, $a_\perp$. We found that $a_\parallel$ had little impact on accuracy (**Table S1**), so we did not consider it in further computations. The social information is the neighbour position with respect to the focal, x_i and y_i, its velocity, $v_{i,x}$ and $v_{i,y}$, and acceleration, $a_{i,x}$ and $a_{i,y}$. Neighbour accelerations had little impact on accuracy

(**Table S1**) and were not used in further computations.

Predictability of the turning side of the focal fish after 1 second improves with the number of neighbours, but with diminishing returns (**Fig. S2**); we chose $n = 25$ neighbors. The tangential acceleration of the focal and the acceleration of the neighbours have a small impact on accuracy of $< 0.1\%$ (**Table S1**) and we thus discarded them in further analysis. In the main text we provide analysis of groups of 100 animals and prediction at 1 s in the future for illustration purposes. Our models predict well a range of futures (**Fig. S3**). Results on how fish interact were found to be similar in computations using 250 ms, 500 ms and 1.5 s in the future (**Fig. S4-S6**) and for groups of 60 or 80 zebrafish (**Fig. S7,8**).

Predictability of the turning side after 1 s is higher for large turning angles than for turning angles close to 0 or 180 degrees (**Fig. S9**). For turning angles of $20-160^\circ$, the interaction network predicted the correct side with an accuracy of 84.4%; up to 87 % for $40 - 100^\circ$ (**Fig. S9**). In contrast, a model using only focal variables failed to obtain a high accuracy and reached only 55 % . The high accuracy of the interaction network shows that the $6 \times 25 = 150$ dimensions capture an important part of the collective dynamics. The instances not predicted may originate from a variety of effects, including higher-order correlations, individuality and non-markovian effects, i.e. history-dependency, be it at short scales or at long scales (internal states or unaccounted behavioural variables, like posture or eye movements). Accuracies were larger the larger the group (**Table S2**), consistent with the idea that interactions lock individuals into social dynamics, less stochastic and less dependent on individual internal states than asocial dynamics.

2.2 Deep attention networks obtain a predictive and analyzable model

The interaction model obtained is too high-dimensional to provide useful insight of animal interactions. The pair interaction subnetwork takes the values of 6 variables as inputs and outputs 128 values (**Fig. 1B**). The aggregation subnetwork first sums up the 25 128-dimensional vectors to give a single vector of 128 components, and then processes it to output a single number, z (**Fig. 1C**). However, the interaction network remains a useful reference for how predictive the behavior is.

To gain insight, we used a deep attention network instead [28, 29]. Like the interaction network, the attention network has a subnetwork to describe the interaction of the focal with each of the n neighbors, except now with a single output (**Fig. 1D**). The aggregation subnetwork is a function weighting differently each neighbor depending on its kinematic parameters and those of the focal (**Fig. 1E**). We found that focal and neighbor speed and neighbor position are the inputs to the aggregation subnetwork with the highest impact in accuracy. We can express the probability that the focal turns to the right after 1 s, p, as $p = 1/1 + \exp(-z)$, where z is the logit that the deep attention network outputs (**Fig. 1F**),

$$z = \sum_{i=1}^{n} \Pi(\alpha, \sigma_i) \frac{W(\alpha^{(\mathrm{w})}, \sigma_i^{(\mathrm{w})})}{\sum_j W(\alpha^{(\mathrm{w})}, \sigma_j^{(\mathrm{w})})}. \tag{1}$$

The pair-interaction subnetwork, $\Pi(\alpha, \sigma_i)$, describes the interaction of the focal and one neighbour i. The aggregation network W gives different weights to the different neighbors i in the aggregation depending on the kinematic parameters of focal $\alpha^{(W)}$ and neighbor relative to focal $\sigma_i^{(W)}$. The subscript indicates that these variables may be different to the ones in the pair-interaction subnetwork.

Since we want the pair-interaction subnetwork, Π, and the aggregation subnetwork, W, to represent the logit of turning after 1 s given a neighbor and a weight representing the importance of that neighbor, respectively, they must differ on several accounts. (i) Π can have any real output, while W must be always positive. (ii) Π must be antisymmetric with respect to reflection on the y-axis, while W must be symmetric. This is because we assume that a neighbour to the right makes the focal go to the right as much as an identical neighbour to the left of the focal makes the focal move to the left. For the aggregation weight W, however, we assume that the importance of the two cases is the same. (iii) The aggregation weights must sum 1. These three conditions are required and we enforced them by: (i) using an exponential as final activation function of W, (ii) antisymmetrizing Π and using symmetric input in W, and (iii) normalising the outputs of W by the sum across all neighbours prior to the integration with the outputs of Π.

The resulting attention network achieves 83.2% accuracy for turns between 20° and 160° and around 85.9% for 30-100° (**Fig S9**). This is slightly less accurate than the interaction network, but the much lower dimensionality of the two subnetworks allows for a detailed analysis.

2.3 The structure of interaction of a pair of animals in a collective

The pair-interaction subnetwork Π is a six-dimensional function. We plotted its output, the logit of the focal fish turning to the right after 1 s, z, as a function of two variables: the angle, θ_i, and the speed of the neighbour, v_i (**Fig. 2**). We fixed the other four variables: focal at median velocity of 3.04 BL/s, focal normal acceleration at $a_\perp = 0$, and neighbor position at $x_i = 7$ BL and $y_i = 1$ BL). At a neighbour velocity above the median (**Fig. 2A**, left; median speed indicated with a horizontal line at 3.04 BL/S), the focal animal is sensitive to the neighbour orientation, with a high probability of turning right (left) after 1 s when the neighbour is moving away from (toward) the focal, resulting in an alignment of the focal to the neighbour. When the neighbour speed is below the median, however, the focal is attracted towards it regardless of the neighbour orientation.

As a contrasting example, consider when the neighbour is closer and slightly in front, at $x_i = 3$ BL and $y_i = 1$ BL (**Fig. 2A**, right). In this case, the focal gets repelled by the neighbour when the neighbor speed is below 3 BL/s.

These two examples illustrate how alignment, attraction and repulsion depend not only on the neighbour location but also on its speed (**Fig. 2B**, similar to **Fig. 2A** but for a 8x8 matrix of subplots, each for a different neighbour position), and on the speed and acceleration of the focal (**Fig. S10-14**).

From this six-dimensional function we can define alignment regions as those where the logit changes sign with neighbour orientation, that is, when focal will turn right (left) if neighbour orients to the right (left) (**Fig. 3**, gray regions. The alignment score (see **Materials and Methods**) measures how sensitive the logit is to neighbour orientation (**Fig. 3A**, gray region; focal speed fixed at median value of 3.04 BL/s and neighbour speed indicated on top of each subplot). The alignment region increases in size and in score with increasing neighbour speed. At high neighbour velocities, strong alignment areas are 2-5 BL behind the focal and 3-5 BL at the sides (**Fig. 3A**, right, darker gray regions). In a region 5-7 BL behind the focal there is a weak orientation score but reversed in sign, with focal turning right (left) when neighbour orients to the left (right), (**Fig. 3A**, pink). This anti-alignment region extends when increasing focal speed, while keeping neighbour speed fixed at the median value of 3.04 BL/s (**Fig. 3B**, pink).

We define attraction (repulsion) regions as those where the logit does not change sign when changing the neighbour angle. Instead, the focal is attracted towards (repelled from) the neighbour's location independently of its orientation. The attraction-repulsion score (see **Materials and Methods**) measures how positive (attraction) or negative (repulsion) is the logit of turning towards the neighbour (**Fig. 3A**). Attraction regions shrink with increasing neighbour speed. They are mainly located to the side at 6-8 BL, extending to the back. Repulsion takes place only when the neighbour speed is below the median speed and neighbours are close to the focal (**Fig. 3A,B**, purple).

However, classifying interactions into only 4 classes is oversimplistic, and a more complete account is captured by the six dimensional pair-interaction function in **Fig. 2**. For example, when the neighbour is at $(x_i, y_i) = (3, 1)$ BL and at high velocity, there is alignment but with a much higher probability of turning left at angles below $\pi/2$ than turning right at angles above $\pi/2$. This asymmetry in angles makes the sensitivity to orientation to the neighbour a mix of alignment and repulsion. We can see the full extent of relative attraction and repulsion zones by plotting the average over neighbour orientation angles for all points in space (**Fig. 3C** for different neighbour speeds and **Fig. S15** for different focal speeds). There is an approximately 5 BL diameter region of relative repulsion around the focal. Regions with a mix of alignment and repulsion (or attraction) are those of alignment in **Fig. 3A** and **Fig. 3B** that overlap with regions of relative repulsion (attraction) in **Fig. 3C** and **Fig. S15**, respectively.

2.4 How information is aggregated: shifting between majorities and minorities

The aggregation subnetwork W outputs the (positive) weight of each neighbor in the aggregation. We found it to depend mainly on 4 variables: focal speed v as the asocial variable and neighbor speed v_i and relative position of neighbour, x_i and y_i. In each subplot of **Fig. 4A**, we give W for

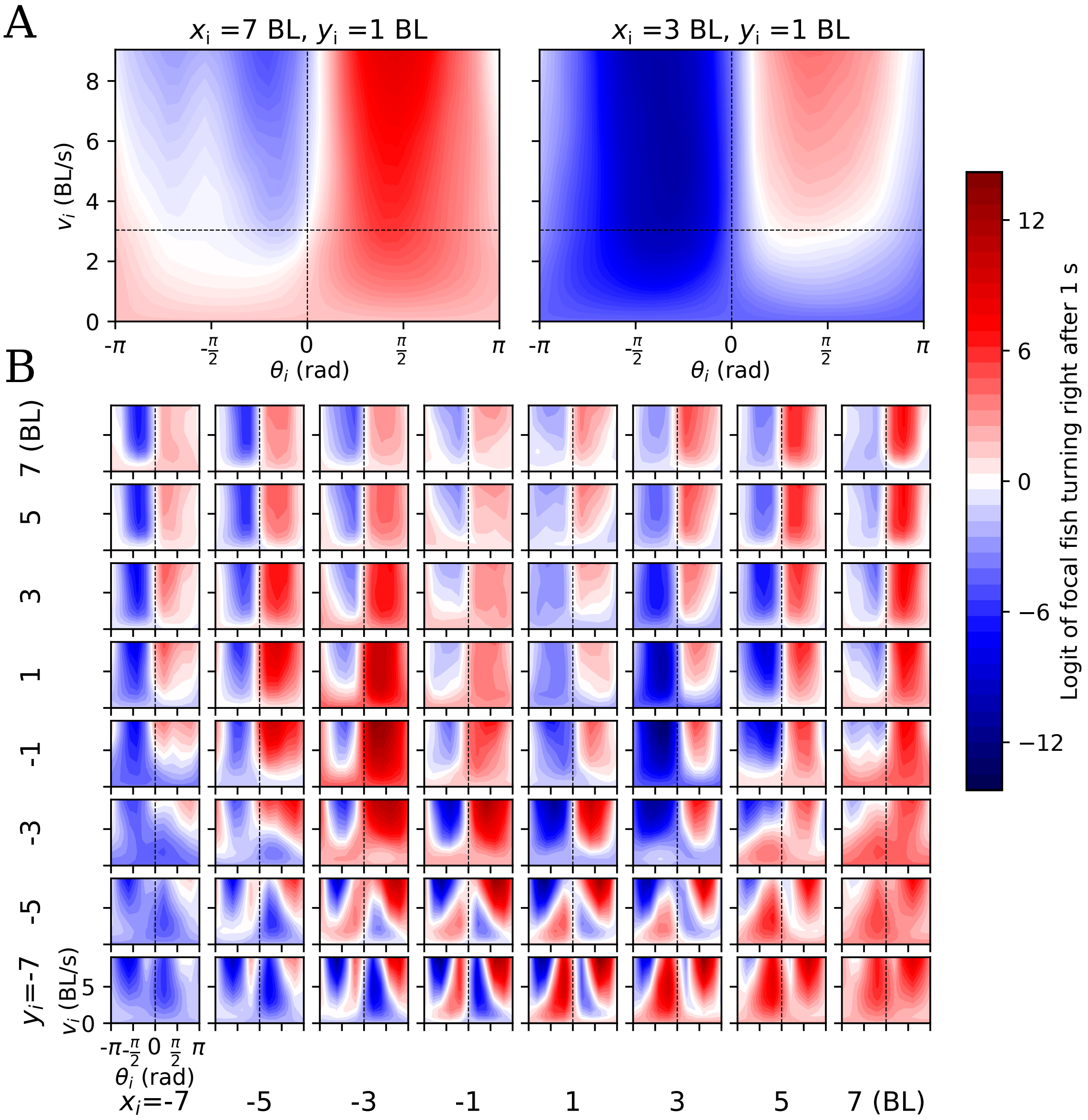

Figure 2: **Properties of interaction between a pair of fish in the collective**. (**A**) Logit z resulting from the pair-interaction subnetwork of the attention network, plotted as a function of the orientation of the neighbour respect to the focal, θ_i, and speed of the neighbour, v_i, for neighbour located at $(x_i, y_i) = (7, 1)$ BL (left) and $(x_i, y_i) = (3, 1)$ BL (right). Focal speed is fixed at median velocity of 3.04 BL/s and focal acceleration at $a_\perp = 0$ BL/s^2. Red colour is evidence that the focal fish will turn right in 1s, while blue is evidence that the focal fish will turn left. Horizontal dashed line highlights the median speed of 3.04 BL/s. **(B)** Same as (A) but for 64 different neighbour positions (x_i, y_i), with x_i and y_i taking values in $(-7, -5, -3, -1, 1, 3, 5, 7)$.

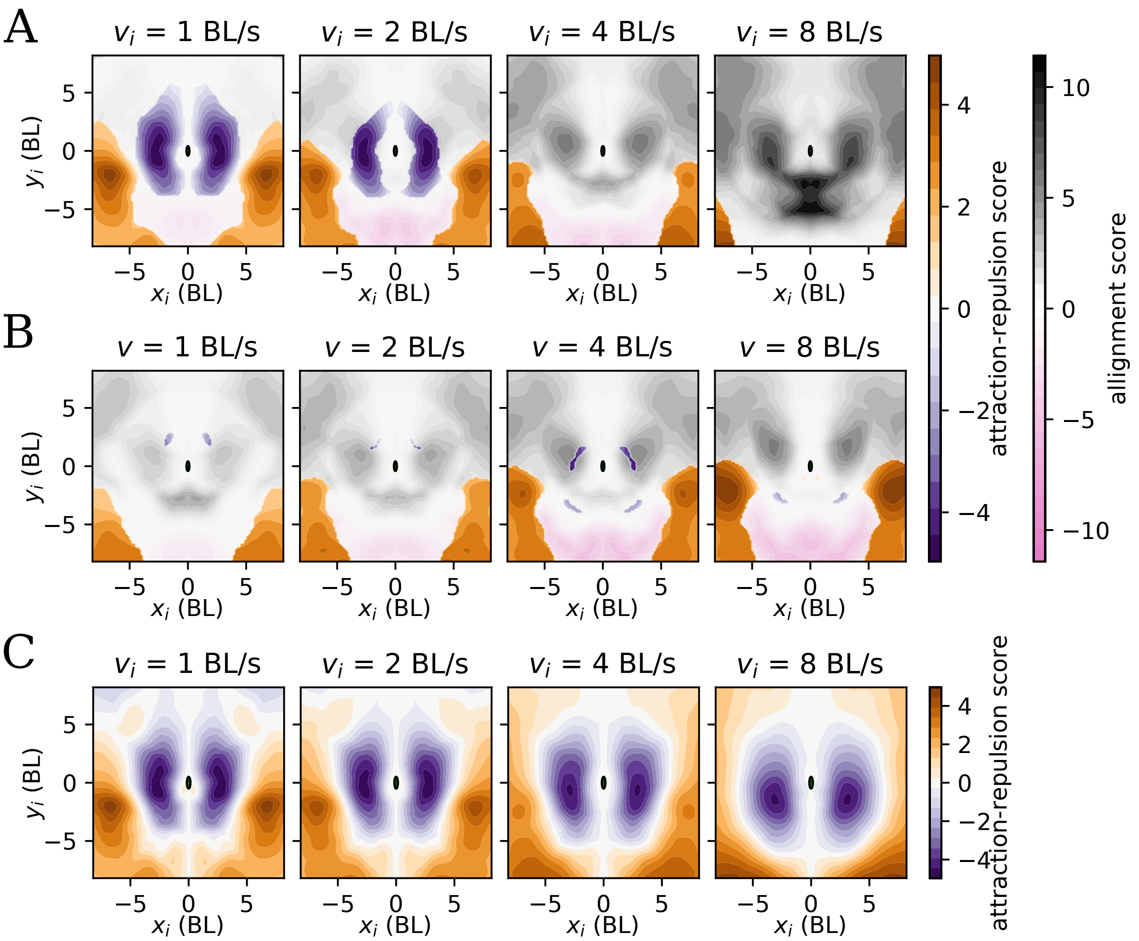

Figure 3: **Alignment, attraction and repulsion zones depend on kinematic parameters of focal and neighbour (A, B)**. Alignment (gray), attraction (orange), repulsion (purple) and anti-alignment (pink) zones. Plotted at four different values of the neighbour speed (1, 2, 4 and 8 BL/s) while keeping focal speed fixed at the median speed 3.04 BL/s **(A)**, and at four different focal speeds (1, 2, 4 and 8 BL/s) while keeping neighbour speed fixed at the median speed 3.04 BL/s **(B)**. Attraction-repulsion score is computed as the average logit across relative orientation angles, θ_i, and it is computed for those regions without change in logit when changing orientation angle. Alignment score measures the range in logit to changes in orientation angle, and it is computed for those regions with change in sign of logit when changing orientation angle (see **Materials and Methods**). It is negative (anti-alignment) when it is anticorrelated with the relative orientation angle (pink). Focal speed fixed at median speed of 3.04 BL/s and focal normal velocity fixed at $a_\perp = 0$. (**C, D**) Relative attraction and repulsion zones. In contrast to (A,B), the logit is averaged over orientation angles also in the regions that were previously of alignment. Regions of alignment in (A,B) mixed with attraction or repulsion show here as relative attraction or repulsion.

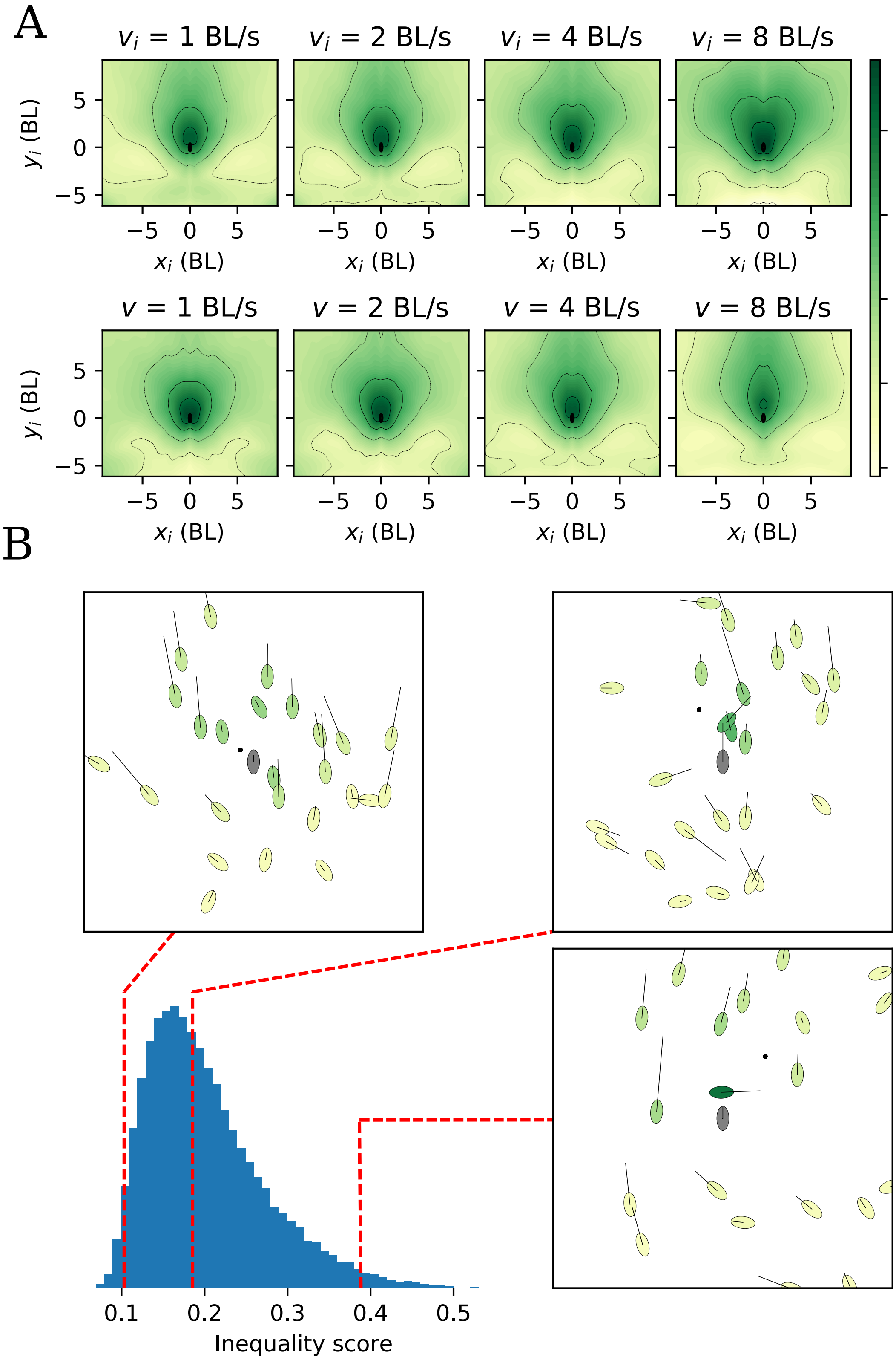

Figure 4: **How a fish aggregates information from neighbours**. **A**. Logarithm of the aggregation weight, $\log(W)$, as a function of neighbour position, x_i and y_i. Top row: focal speed fixed at 3.04 BL/s and each subplot corresponding to different neighbour speeds marked on top of each. Bottom row: same as top row but for fixed neighbour speed at 3.04 BL/s and different focal speeds. **B**. Distribution of inequality score (value of highest weight). Three example frames with each neighbour colored with its weight in the aggregation. Focal animal is indicated in gray color, with a horizontal line proportional to the normal acceleration to either left or right and a small dot in its frontal positions indicating the focal position 1 second into the future.

different neighbour positions, keeping neighbour and focal speed constant. Generally, W is higher for neighbours that are closer to the focal, and lower for neighbours behind the focal. In the upper row of **Fig. 4A** all subplots have the same focal speed at the median velocity of 3.04 BL/s, and each indicates the neighbour speed on top, with values $v_i = 1, 2, 4$ and 8 BL/s. We see how W increases with neighbour speed for most neighbour positions, implying that faster neighbours carry more weight in the aggregation. This is more pronounced close by and to the side. In the lower row of **Fig. 4A**, all subplots have the same neighbor speed at the median value and each one indicates above the focal speed. We see how the mass of W increasingly shifts towards the front the faster the focal fish moves. Adding other variables to the attention marginally improves accuracy, and still further insight is gained. When the neighbour orientation angle is added, higher values of the weight W are obtained in positions leading to an immediate collision (**Fig. S16**). The final impact of each neighbour on the probability of the focal turning right is given by its weight relative to the weights of the rest of neighbours, $W(\alpha^{(\mathrm{w})}, \sigma_i^{(\mathrm{w})}) / \sum_j W(\alpha^{(\mathrm{w})}, \sigma_j^{(\mathrm{w})})$. For example, if all neighbors are assigned the same weight by W, no matter how large or small this value is, the final logit is the average of the individual logits, $z = \frac{1}{n} \sum_{i=1}^{n} \Pi(\alpha, \sigma_i)$. If one of the neighbours has a higher (lower) value of W, its importance in the integration increases (decreases), while the importance of the other neighbours decreases (increases).

The highest relative weight among all neighbours can be used to quantify the inequality of relative weights at any point in time. We find a wide distribution of values of this inequality score (**Fig. 4B**). In some cases, all neighbours have similar weights, and the inequality score is small (**Fig. 4B**, upper left). Most often, a subgroup of neighbours is weighed more, the most common inequality score being just below 0.2 (**Fig. 4B**, upper right). The distribution of the inequality score has a long tail, making high inequality disproportionately common. In some cases, one neighbour overweighs the others by far (**Fig. 4B**, lower right), up to the combined weight of all others.

3 Discussion

Our results show that animals in collectives can use an aggregation rule that naturally allows individuals to shift from simple averaging to following a single individual. This helps to close the gap between models using averages [11, 12, 15, 16] and few individuals [22, 23]. Also, it opens the door to study how animals match interactions among themselves to the knowledge distribution in the group [32, 26]. Note that our models and its source code can be reused for any species.

From the pairwise interaction in the collective, we could extract attraction, repulsion and alignment as approximate notions [12, 13]. Usually, these interaction classes are defined only in terms of relative position of neighbour. However, we found them to exist in a 6-dimensional space. This translates into these classes also depending on speeds of focal and neighbour, focal acceleration and relative orientation between the two fish. Moreover, the three classes are not cleanly separated as alignment regions are mixed with attraction or repulsion.

As a strategy to extract the relevant variables for behavior, we have required them to predict future behavior (like e.g. [33, 34, 30, 16]). This approach has the additional advantage of automatically generating labelled data for supervised training of networks. It can be enriched, at the cost of increasing the model dimensionality, with more information about behavioral history, possible internal variables (parametrized, for example, by time of day or more direct internal measurements), explicit dynamics and posture using reduced variables [35, 36, 37, 38]. A second requirement for our models was that they should work for data not used to obtain the model. These two requirements are standard in machine learning, but not in the study of collective animal behavior.

Our results illustrate how modular deep networks enable flexible data-driven modelling without losing insight. Each module is flexible, with tens of thousands of parameters, but implements a function with low dimensionality in the number of inputs and outputs. Combinations of modules [39, 40], two types in the attention network, achieve high compositional complexity that adds flexibility without losing insight.

4 Methods

4.1 Data availability

60- and 100-fish as well as the new 80-fish videos can be found at www.idtracker.ai. Code is free and open source software (https://github.com/fjhheras/fishandra)

4.2 Animal rearing and handling

Zebrafish, *D. rerio*, of the wild-type TU strain were raised by the Champalimaud Foundation Fish Platform, according to methods in [41]. Experimental procedures were approved by the Champalimaud Foundation Ethics Committee and the Portuguese Direcção Geral Veterinária in accordance to to the European Directive 2010/63/EU. Handling procedures were as in [27]. We used juveniles of 31-33 days post fertilization.

4.3 Videos and tracking

We used 6 videos of 60 and 100 freely swimming juvenile zebrafish from [27], and 3 new videos of 80 juveniles. The camera had a frame rate of 32 fps and 20 Mpx of definition. We obtained all the fish trajectories using *idtracker.ai* (**Fig S1**) with an accuracy of 99.95 (mean) ±0.01% (std) [27].

4.4 Preprocessing

We interpolated linearly the very small holes in the tracked trajectories (0.027% for 100-fish videos). We normalized trajectories, by translation (center of arena at (0,0)) and scaling (radius of the arena at 1). To reduce noise while preventing contamination by any future information, we smoothed the trajectories using a 5-frame half-Gaussian kernel with $\sigma = 1$ frame. We obtained velocity and acceleration by finite differences, using only current and past frames. To avoid direct border effects, we removed datapoints where the focal fish is further away from the center than 80% of the radius. Each video was divided in three parts, to obtain the training, validation and test datasets (97%/2%/1%).

In each video frame, for each individual, we found the n nearest neighbours ($\mathcal{I}$). We then obtained (i) velocity and acceleration of the focal fish, (ii) relative position, absolute velocity and absolute acceleration of the closest n neighbours, (iii) whether the focal fish has turned right or left after N_f frames in the future.

4.5 Deep networks

We implemented the Deep Networks using TensorFlow through its Python API [42]. We solved the following classification task: Given dynamical properties of a focal fish and its n closest neighbours, does the focal fish turn right or left after 1s? Asocial information is the set of speed and normal and tangential acceleration of the focal,

$$\alpha = \{v, a_\perp, a_\parallel \dots\}. \tag{2}$$

Social information from a neighbour i is its location, velocity and acceleration

$$\sigma_i = \{x_i, y_i, v_i, \theta_i, a_{\mathrm{x},i}, a_{\mathrm{y},i}, \dots\}, \tag{3}$$

whose coordinates we calculate in an instantaneous frame of reference that is not moving, which is centered in the focal fish and whose y-axis is co-lineal with the focal fish velocity. Note that v_i is the absolute speed, while (x_i, y_i) is the relative position of the neighbour, rotated to the frame of reference. In each network, we first obtain the logits z, and then the probabilities by using a logistic function $p = 1/(1 + e^{-z})$.

4.5.1 Interaction network

In the interaction network [31], given asocial (α) and social ($\{\sigma_i, i \in \mathcal{I}\}$) information, the logit of turning right is calculated as

$$z = I(\alpha, \{\sigma_i\}) = \Gamma\left(\sum_{i\in\mathcal{I}} \Pi_{\mathrm{I}}(\alpha, \sigma_i)\right). \tag{4}$$

The function Π_{I} is the pair-interaction subnetwork. We modelled it using a fully-connected network with 3 hidden layers of 128 neurons each, plus a readout layer of 128 neurons. There are rectified linear unit (ReLU, [43]) nonlinearities after each hidden layer (but not after the readout). The outputs of Π_{I} for different neighbours are summed together and transformed by a second function, Γ. We modelled Γ as a fully-connected layer with one hidden layer of 128 neurons, plus a one-neuron readout layer. There are ReLU nonlinearities preceding the whole network and after each hidden layer (but not after the one-neuron readout layer).

To effectively multiply available data by n, we considered all neighbours to be equal. Equivalently, there is symmetry with respect to exchange of neighbour labels. We did not observe any turning side preference. Therefore, to effectively multiply available data by 2, we forced the network to be antisymmetrical with respect to a reflection along the body axis by antisymmetrization of I,

$$z = I(\alpha, \{\sigma_i\}) - I(\alpha^*, \{\sigma_i^*\}), \tag{5}$$

where the star superscript represents a reflection along the longitudinal axis of the body, calculated by switching the sign of all x components.

4.5.2 Attention network

Eq. [1] in main text can be rewritten using a notation that compares directly with Eq. [4] in Methods as

$$z = A(\alpha, \{\sigma_i\}) = \sum_{i \in \mathcal{I}} \Pi_{\mathrm{A}}(\alpha, \sigma_i) \frac{W(\alpha, \sigma_i)}{\sum_k W(\alpha, \sigma_i)}. \tag{6}$$

The function Π_{A} captures the effect of pairwise interactions. It has the same structure as Π_{I} except that its readout layer has only one neuron, and that we antisymmetrise it. W is an attention layer, weighting the logits of the different neighbours. W has the same structure as Π_{A}, except that it accepts as input a y-axis-reflection-invariant subset of the asocial and social variables, and that there is an exponential function after the single-neuron readout signal.

4.5.3 Loss

Following standard procedures in binary classification, when training the network to estimate the probability p_i of turning right, we minimised the cross-entropy loss, [43]

$$\mathcal{L} = -\frac{1}{N_b} \sum_{i=1}^{N_b} \log(p_i^*). \tag{7}$$

summed along N_b data points in the minibatch and where p_i^* is the probability given by the network to the actual turn. When the network predicts a right turn with probability p_i, $p_i^* = p_i$ if the actual turn was to the right, and $p_i^* = 1 - p_i$ if the actual turn was to the left. We minimise loss using Adam [43]. We stopped training if validation loss did not reach a local minimum for 10 epochs and did increase 25% from the minimum, or after 100 training epochs. In the attention network, we annealed learning rate from 10^{-4} to 10^{-5}, using a batch size of 500. In the attention network, we annealed learning rate from 5×10^{-5} to 10^{-5} and trained with a batch size of 200. Dropout [43] did not improve accuracy.

4.6 Attraction-repulsion and alignment scores

Attraction-repulsion score is obtained as

$$\mathrm{sign}(x) \langle z_i \rangle_{\theta_i \in [0, 2\pi)}, \tag{8}$$

with attraction (repulsion) when the score is positive (negative). Alignment score is obtained as

$$\max_{\theta_i \in [0,2\pi)} \{z_i \, \mathrm{sign}(\theta_i)\} - \max_{\theta_i \in [0,2\pi)} \{-z_i \, \mathrm{sign}(\theta_i)\}. \tag{9}$$

Acknowledgements

We thank Kristin Branson, Jimmy Liao and the Collective Behavior Lab at Champalimaud for feedback and Ana Catarina Certal for animal care. This work was supported by Fundação para a Ciência e Tecnologia PTDC/NEU-SCC/0948/2014 (to GGdP and contract to FJHH), Champalimaud Foundation (to GGdP) and GGdP work at KITP Santa Barbara by NSF Grant No. PHY-1748958, NIH Grant No. R25GM067110, and the Gordon and Betty Moore Foundation Grant No. 2919.01.

Contributions

GGdP and FJHH designed study, decided on modelling and analyzed data, FJHH wrote the code, RH and FR-F built experimental setup, recorded and tracked videos, GGdP supervised research and GGdP and FJHH wrote the manuscript.

References

[1] Nicolas De Condorcet. *Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix*. L'imprimerie royale, 1785.

[2] Francis Galton. Vox populi (the wisdom of crowds). *Nature*, 75(7):450–451, 1907.

[3] Ludmila I Kuncheva, Christopher J Whitaker, Catherine A Shipp, and Robert PW Duin. Limits on the majority vote accuracy in classifier fusion. *Pattern Analysis & Applications*, 6(1):22–31, 2003.

[4] Andres Laan, Gabriel Madirolas, and Gonzalo G. de Polavieja. Rescuing collective wisdom when the average group opinion is wrong. *Frontiers in Robotics and AI*, 4:56, 2017.

[5] Philip J Boland. Majority systems and the condorcet jury theorem. *The Statistician*, pages 181–189, 1989.

[6] Robert Edward Goodin. *Reflective democracy*. Oxford University Press on Demand, 2003.

[7] Melissa Schwartzberg. Epistemic democracy and its challenges. *Annual review of political science*, 18:187–203, 2015.

[8] Robert E Goodin and Kai Spiekermann. *An epistemic theory of democracy*. Oxford University Press, 2018.

[9] Iain D Couzin, Jens Krause, Nigel R Franks, and Simon A Levin. Effective leadership and decision-making in animal groups on the move. *Nature*, 433(7025):513, 2005.

[10] Aviram Gelblum, Itai Pinkoviezky, Ehud Fonio, Abhijit Ghosh, Nir Gov, and Ofer Feinerman. Ant groups optimally amplify the effect of transiently informed individuals. *Nature communications*, 6:7729, 2015.

[11] Tamás Vicsek, András Czirók, Eshel Ben-Jacob, Inon Cohen, and Ofer Shochet. Novel type of phase transition in a system of self-driven particles. *Physical review letters*, 75(6):1226, 1995.

[12] Iain D Couzin, Jens Krause, Richard James, Graeme D Ruxton, and Nigel R Franks. Collective memory and spatial sorting in animal groups. *Journal of theoretical biology*, 218(1):1–11, 2002.

[13] Hugues Chaté, Francesco Ginelli, Guillaume Grégoire, Fernando Peruani, and Franck Raynaud. Modeling collective motion: variations on the vicsek model. *The European Physical Journal B*, 64(3-4):451–456, 2008.

[14] Tamás Vicsek and Anna Zafeiris. Collective motion. *Physics Reports*, 517(3-4):71–140, 2012.

[15] Jacques Gautrais, Francesco Ginelli, Richard Fournier, Stéphane Blanco, Marc Soria, Hugues Chaté, and Guy Theraulaz. Deciphering interactions in moving animal groups. *Plos computational biology*, 8(9):e1002678, 2012.

[16] Roy Harpaz, Gašper Tkačik, and Elad Schneidman. Discrete modes of social information processing predict individual behavior of fish in a group. *Proceedings of the National Academy of Sciences*, 114(38):10149–10154, 2017.

[17] Ashley JW Ward, David JT Sumpter, Iain D Couzin, Paul JB Hart, and Jens Krause. Quorum decision-making facilitates information transfer in fish shoals. *Proceedings of the National Academy of Sciences*, 105(19):6948–6953, 2008.

[18] David JT Sumpter, Jens Krause, Richard James, Iain D Couzin, and Ashley JW Ward. Consensus decision making by fish. *Current Biology*, 18(22):1773–1777, 2008.

[19] Alfonso Pérez-Escudero and Gonzalo G de Polavieja. Collective animal behavior from Bayesian estimation and probability matching. *PLoS computational biology*, 7(11):e1002282, nov 2011.

[20] Sara Arganda, Alfonso Pérez-Escudero, and Gonzalo G de Polavieja. A common rule for decision making in animal collectives across species. *Proceedings of the National Academy of Sciences of the United States of America*, 109(50):20508–13, dec 2012.

[21] Robert C Hinz and Gonzalo G de Polavieja. Ontogeny of collective behavior reveals a simple attraction rule. *Proceedings of the National Academy of Sciences of the United States of America*, 114(9):2295–2300, feb 2017.

[22] Steven L Lima. Back to the basics of anti-predatory vigilance: the group-size effect. *Animal Behaviour*, 49(1):11–20, 1995.

[23] Gilbert Roberts. Why individual vigilance declines as group size increases. *Animal Behaviour*, 51(5):1077–1086, 1996.

[24] Larissa Conradt and Timothy J Roper. Group decision-making in animals. *Nature*, 421(6919):155, 2003.

[25] Albert B Kao, Noam Miller, Colin Torney, Andrew Hartnett, and Iain D Couzin. Collective learning and optimal consensus decisions in social animal groups. *PLoS computational biology*, 10(8):e1003762, 2014.

[26] James AR Marshall, Gavin Brown, and Andrew N Radford. Individual confidence-weighting and group decision-making. *Trends in ecology & evolution*, 32(9):636–645, 2017.

[27] Francisco Romero-Ferrero, Mattia G. Bergomi, Robert Hinz, Francisco J. H. Heras, and Gonzalo G. de Polavieja. idtracker.ai: Tracking all individuals in large collectives of unmarked animals. *arXiv preprint*, abs/1803.04351, 2018.

[28] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. Neural machine translation by jointly learning to align and translate. *arXiv preprint*, abs/1409.0473, 2014.

[29] Kelvin Xu, Jimmy Ba, Ryan Kiros, Kyunghyun Cho, Aaron C. Courville, Ruslan Salakhutdinov, Richard S. Zemel, and Yoshua Bengio. Show, attend and tell: Neural image caption generation with visual attention. *arXiv preprint*, abs/1502.03044, 2015.

[30] Yedid Hoshen. Vain: Attentional multi-agent predictive modeling. In *Advances in Neural Information Processing Systems*, pages 2701–2711, 2017.

[31] Peter W. Battaglia, Razvan Pascanu, Matthew Lai, Danilo Jimenez Rezende, and Koray Kavukcuoglu. Interaction networks for learning about objects, relations and physics. *arXiv preprint*, abs/1612.00222, 2016.

[32] Andress Laan, Raul Gil de Sagredo, and Gonzalo G de Polavieja. Signatures of optimal control in pairs of schooling zebrafish. *Proceedings. Biological sciences*, 284(1852):20170224, apr 2017.

[33] Eyrun Eyjolfsdottir, Kristin Branson, Yisong Yue, and Pietro Perona. Learning recurrent representations for hierarchical behavior modeling. *arXiv preprint arXiv:1611.00094*, 2016.

[34] Federico Bartoli, Giuseppe Lisanti, Lamberto Ballan, and Alberto Del Bimbo. Context-aware trajectory prediction. *arXiv preprint arXiv:1705.02503*, 2017.

[35] Masato Nakajima, Seiichi Uchida, Akihiro Mori, Ryo Kurazume, Rin-ichiro Taniguchi, Tsutomu Hasegawa, and Hiroaki Sakoe. Motion prediction based on eigen-gestures. In *Proc. of the 1st First Korea-Japan Joint Workshop on Pattern Recognition*, 2006.

[36] Matei T Ciocarlie, Corey Goldfeder, and Peter K Allen. Dimensionality reduction for hand-independent dexterous robotic grasping. In *IROS*, volume 7, pages 3270–3275, 2007.

[37] Greg J Stephens, Bethany Johnson-Kerner, William Bialek, and William S Ryu. Dimensionality and dynamics in the behavior of c. elegans. *PLoS computational biology*, 4(4):e1000028, 2008.

[38] Gordon J Berman, Daniel M Choi, William Bialek, and Joshua W Shaevitz. Mapping the stereotyped behaviour of freely moving fruit flies. *Journal of The Royal Society Interface*, 11(99):20140672, 2014.

[39] Roger Grosse, Ruslan R Salakhutdinov, William T Freeman, and Joshua B Tenenbaum. Exploiting compositionality to explore a large space of model structures. *arXiv preprint arXiv:1210.4856*, 2012.

[40] Scott Reed and Nando de Freitas. Neural programmer-interpreters. In *International Conference on Learning Representations (ICLR)*, 2016.

[41] Sandra Martins, Joana F Monteiro, Maria Vito, David Weintraub, Joana Almeida, and Ana Catarina Certal. Toward an integrated zebrafish health management program supporting cancer and neuroscience research. *Zebrafish*, 13(S1):S–47, 2016.

[42] Martín Abadi, Paul Barham, Jianmin Chen, Zhifeng Chen, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Geoffrey Irving, Michael Isard, et al. Tensorflow: a system for large-scale machine learning. In *OSDI*, volume 16, pages 265–283, 2016.

[43] Ian Goodfellow, Yoshua Bengio, Aaron Courville, and Yoshua Bengio. *Deep learning*. MIT press Cambridge, 2016.

Table S1: **Predictability of large turns (test data) when using different variables.** 25 neighbours, interaction network, best of two runs. A good trade-off between complexity and accuracy is using neighbour speed, focal speed and focal normal acceleration.

		Neighbour		
		a_i	v_i	$v_i + a_i$
Focal	a	81.6%	84.1%	84.2%
	v	80.7%	83.9%	83.7%
	$v + a$	81.5%	**84.5%**	**84.5%**
	$v + a_t$	80.9%	83.7%	83.8%
	$v + a_n$	81.5%	**84.5%**	**84.4%**

Table S2: **Predictability of large turns (test data) for videos of different number of animals.** 25 neighbours, best of two runs.

	Interaction	Attention
100 individuals	84.4%	83.2%
80 individuals	77.1%	76.4%
60 individuals	73.4%	72.2%

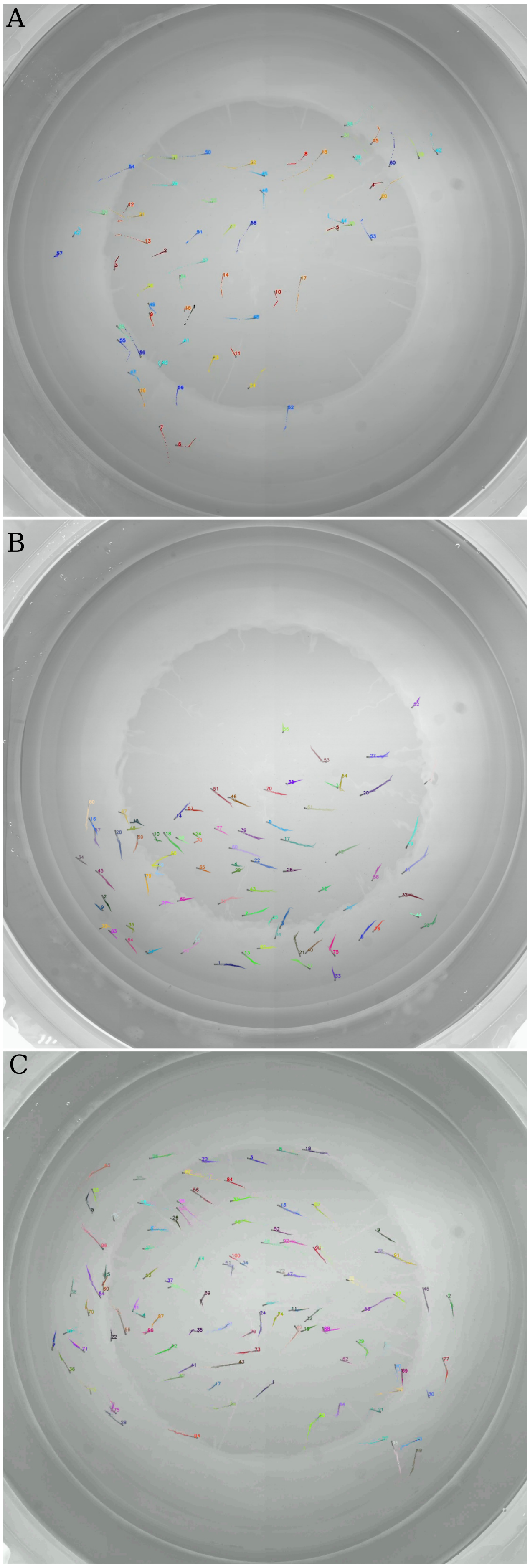

Figure S1: **Example video frames.** Each fish has been individually tracked by identification using idtracker.ai [27]. **A.** 60-fish video. **B.** 80-fish video. **C.** 100-fish video.

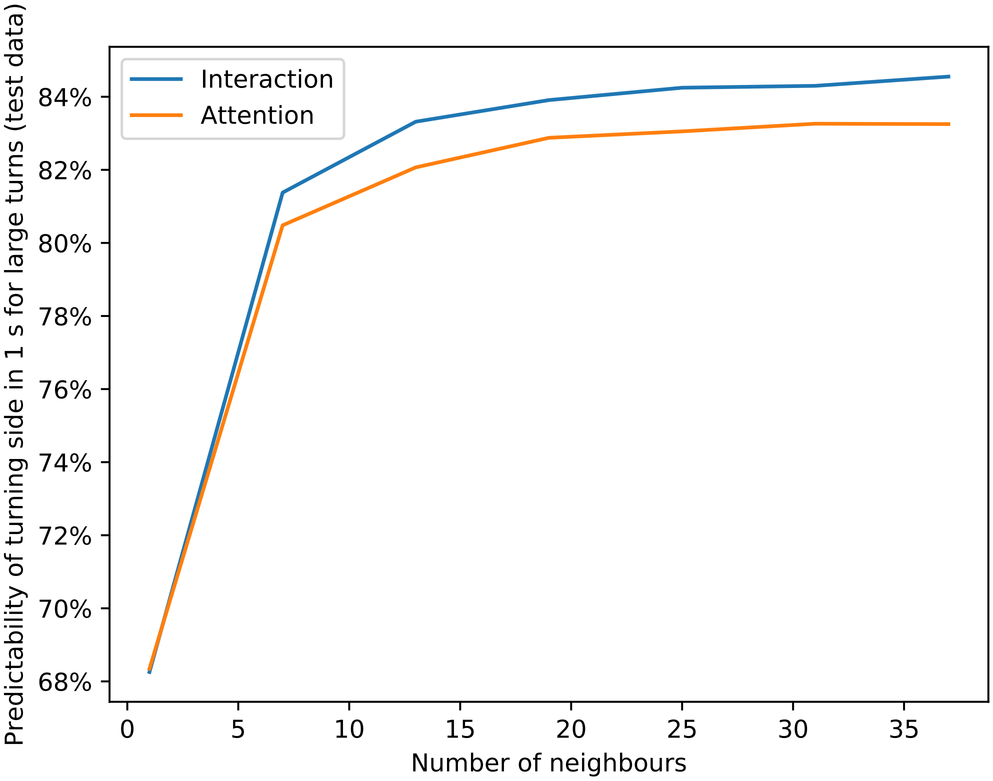

Figure S2: **Predictability of large turns (test set) as a function of number of closest neighbours.** Both the interaction network (blue) and the attention network (orange) improve in accuracy with the number of neighbours, and then plateau after approx. 20 neighbours.

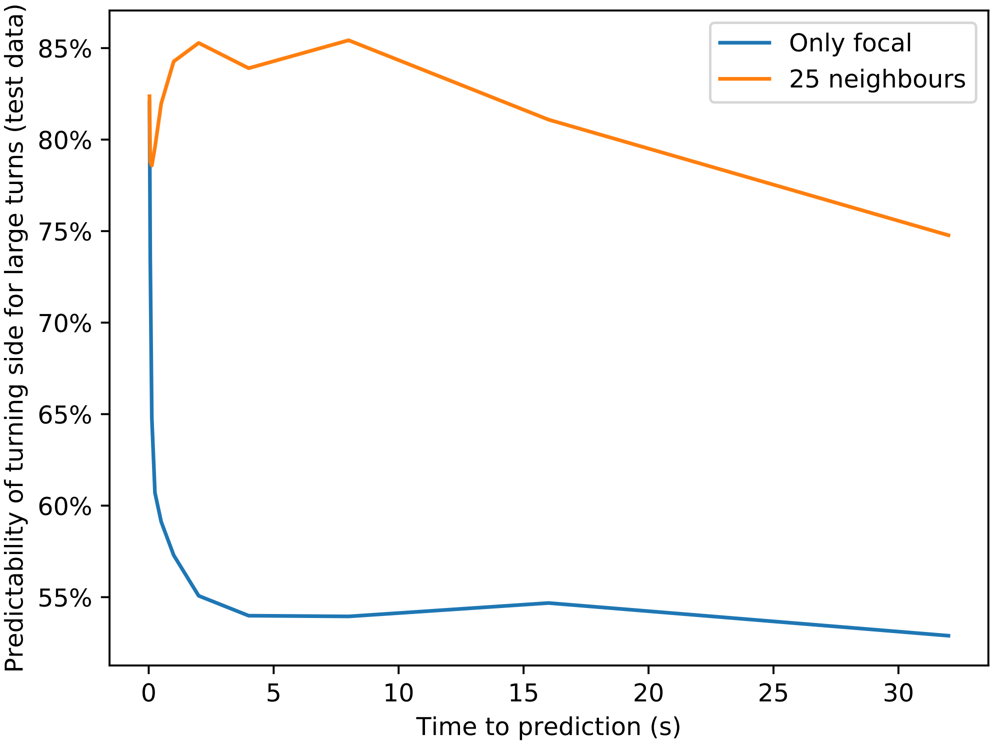

Figure S3: **Predictability (test set) for different times to prediction.** The prediction from an interaction model with 25 neighbours (orange) and from a model that is blind to any social information (blue). Predictability for immediate futures (<100 ms) is high for both models, because of correlations in the acceleration. Predictability with 25 neighbours has a local minimum for futures of 250 ms, it has a broad maximum when predicting futures between 1 and 10 s, and then it drops slowly when predicting more distant futures.

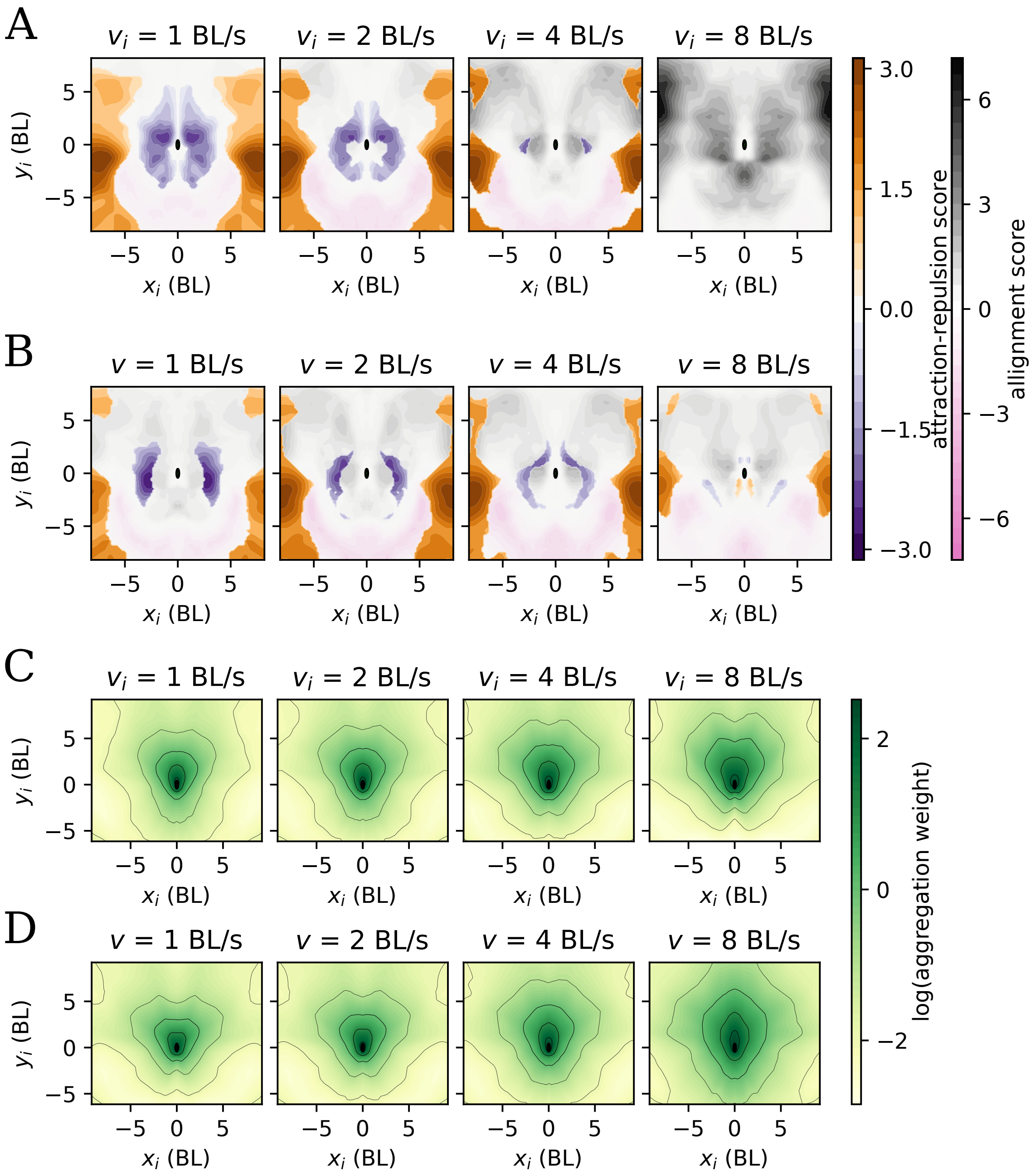

Figure S4: **Pair interaction and aggregation in 250 ms predictions**. **A** Same as Figure 3A. **B** Same as Figure 3B **C** Same as Figure 4A. Note how high-attention areas are closer to the focal fish. **D** Same as Figure 4B

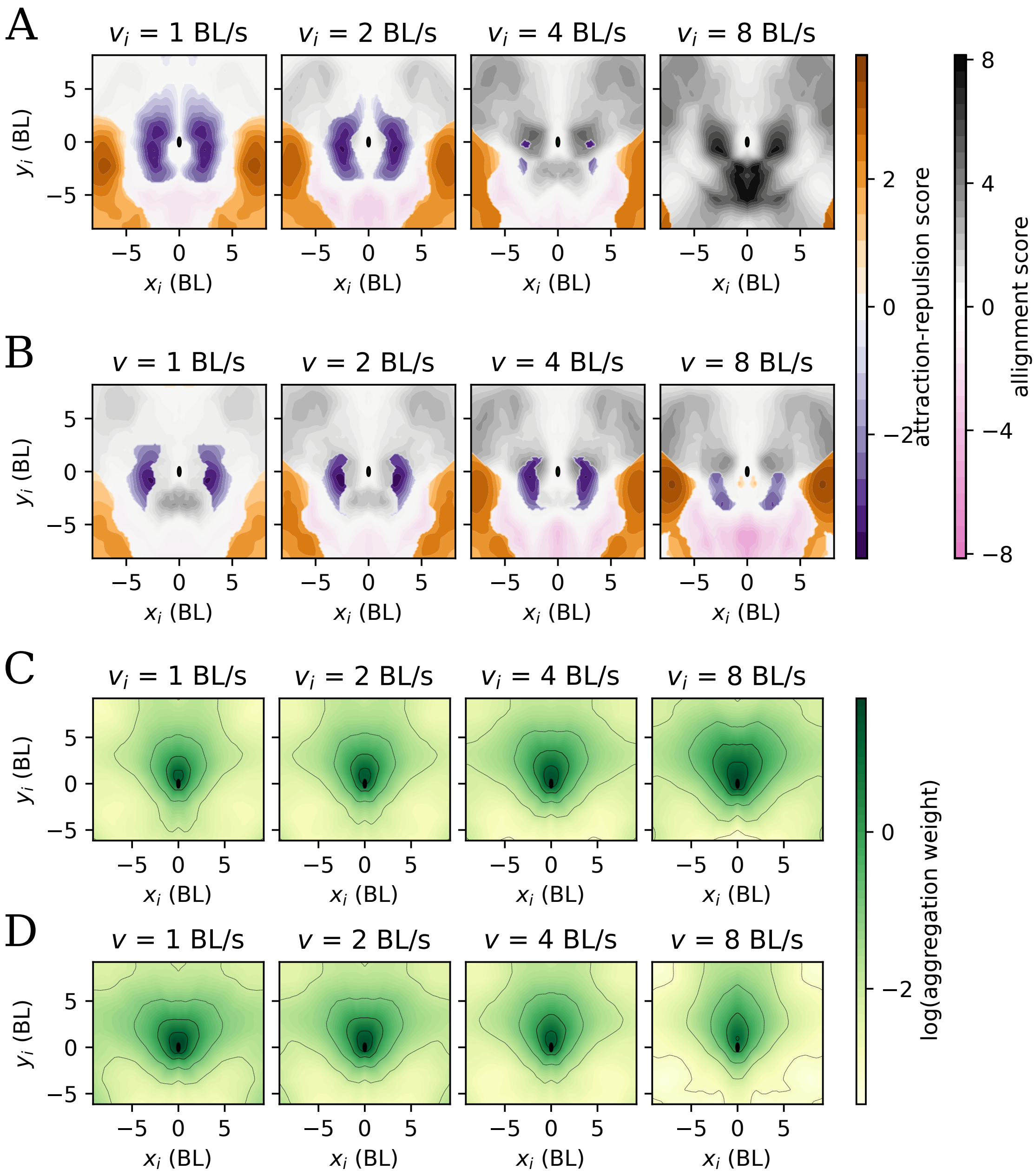

Figure S5: **Pair interaction and aggregation in 500 ms predictions**. **A** Same as Figure 3A. **B** Same as Figure 3B **C** Same as Figure 4A. **D** Same as Figure 4B

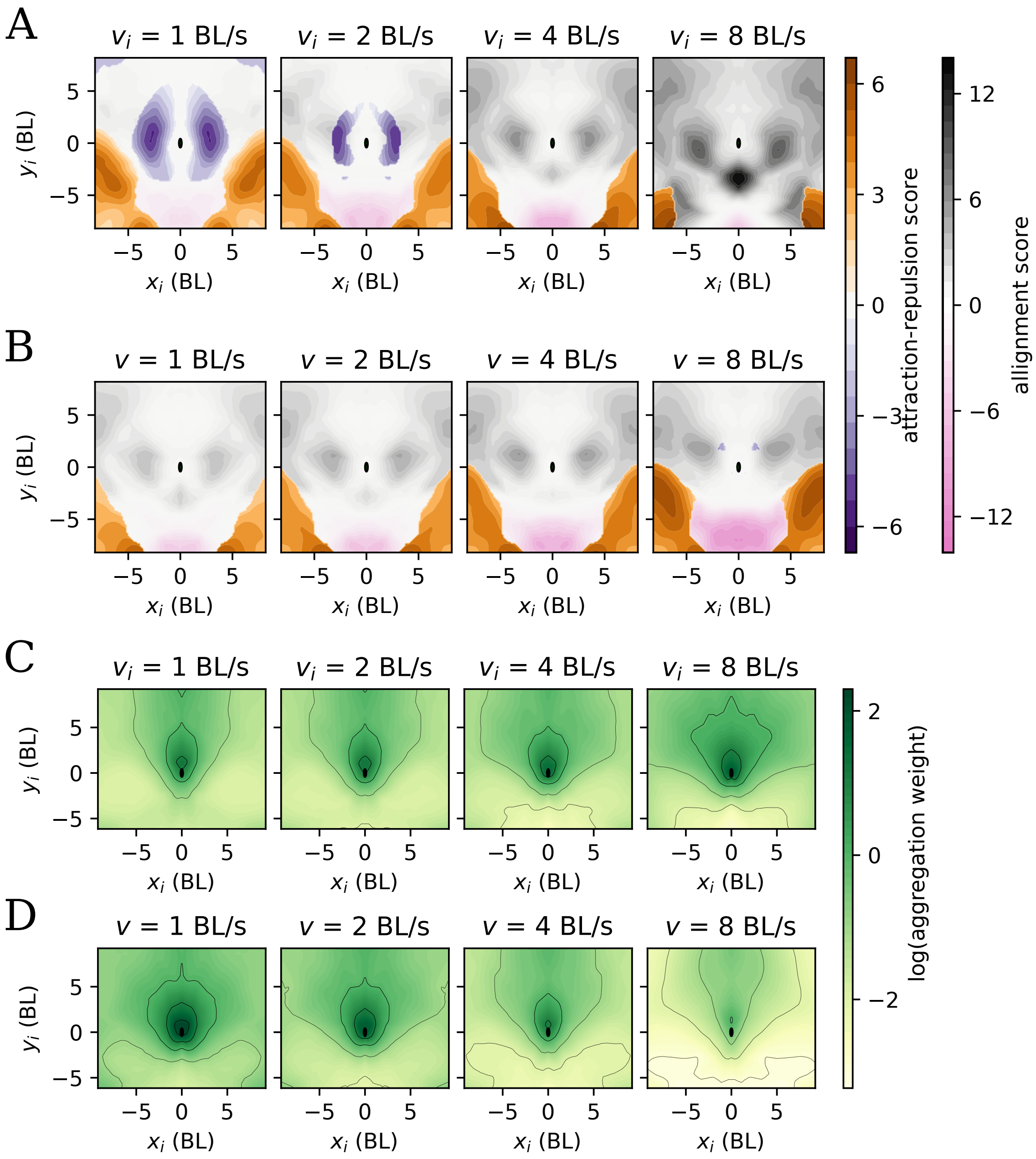

Figure S6: **Pair interaction and aggregation in 1500 ms predictions**. **A** Same as Figure 3A. **B** Same as Figure 3B **C** Same as Figure 4A. Note how high-attention areas are closer to the front. **D** Same as Figure 4B

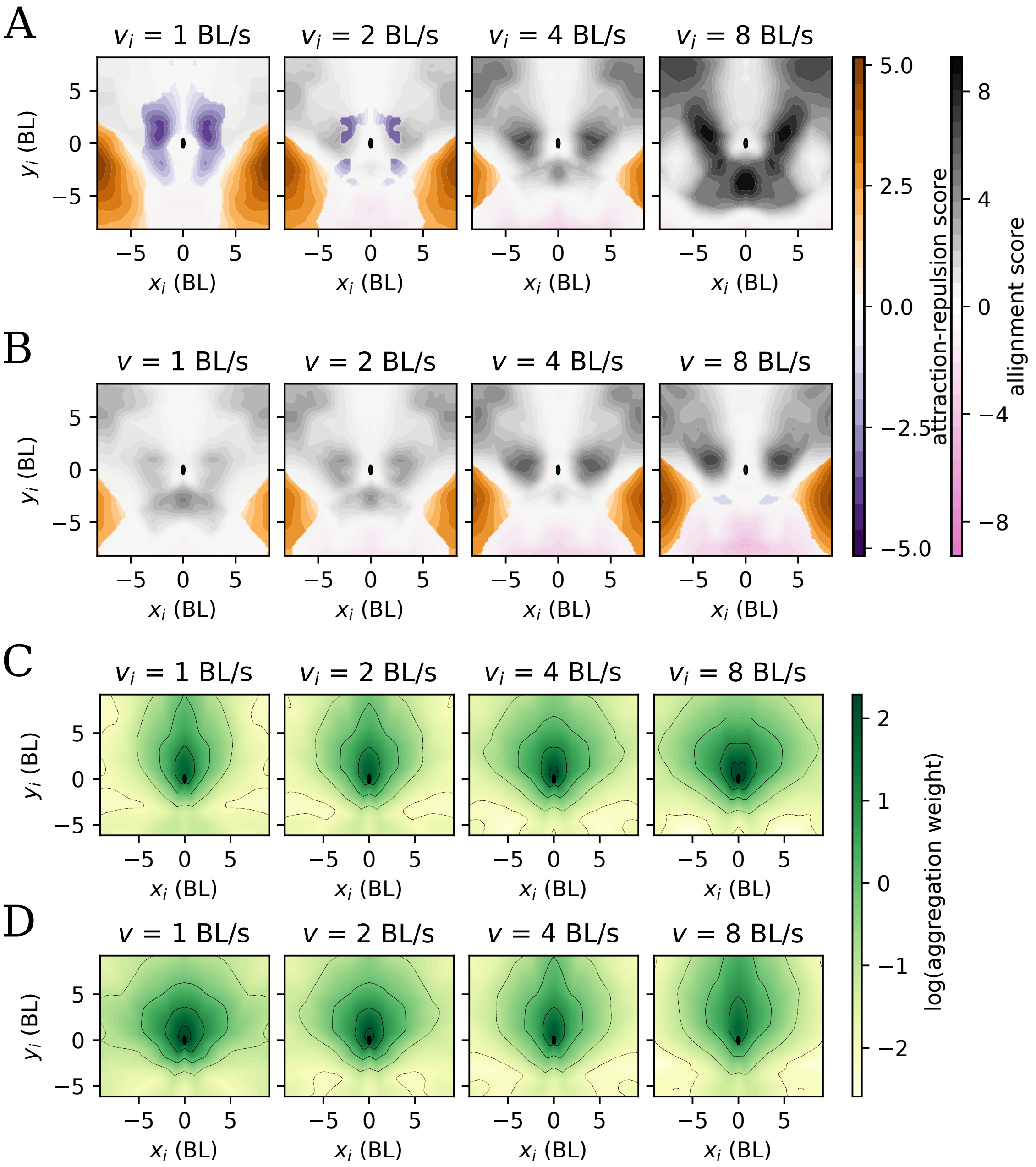

Figure S7: **Pair interaction and aggregation, obtained from 80-fish videos**. **A** Same as Figure 3A. **B** Same as Figure 3B **C** Same as Figure 4A. **D** Same as Figure 4B

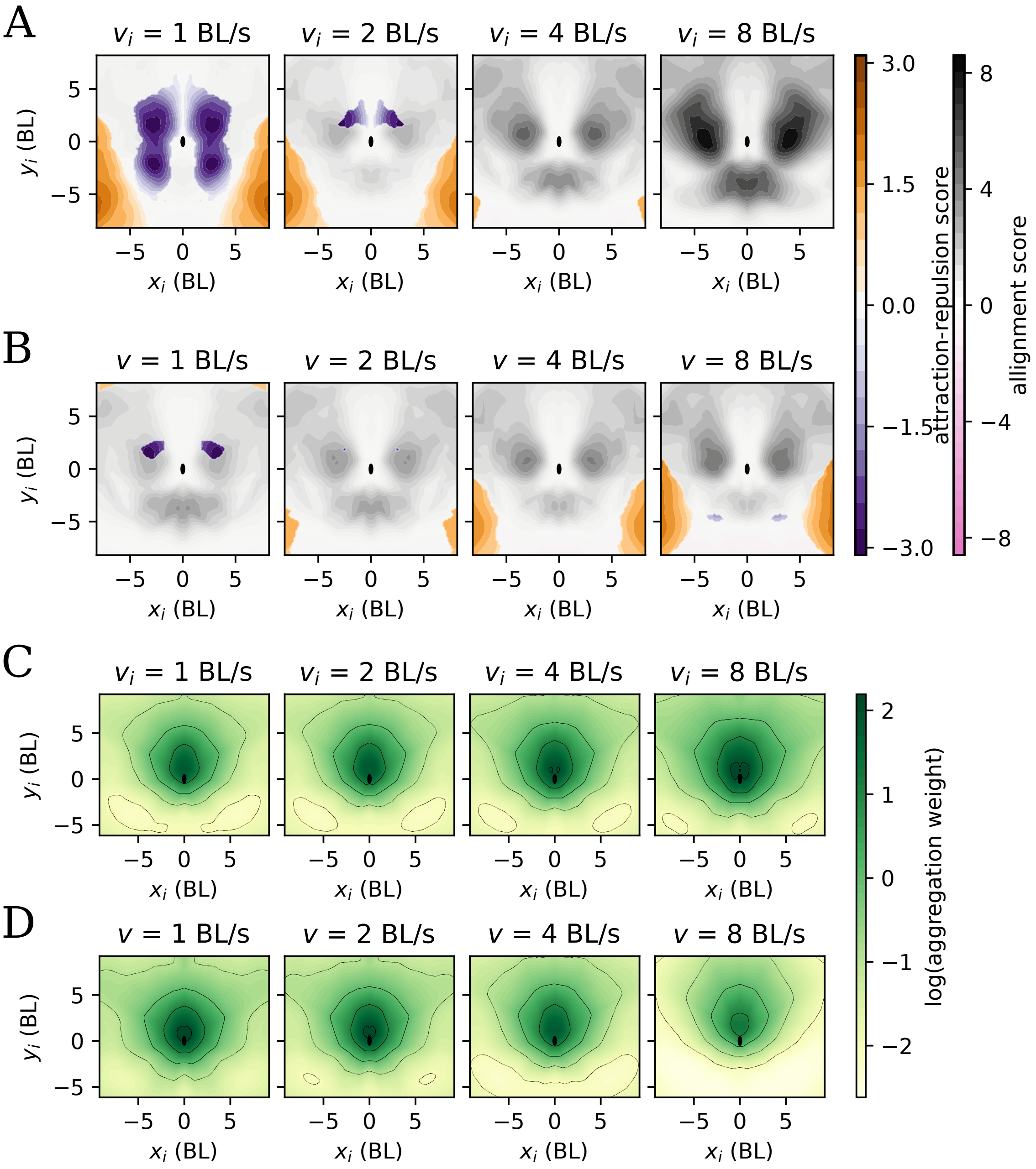

Figure S8: **Pair interaction and aggregation, obtained from 60-fish videos**. **A** Same as Figure 3A. The most conspicuous difference with Figure 3A is the weakening of anti-alignment **B** Same as Figure 3B **C** Same as Figure 4A. **D** Same as Figure 4B. Note the comparatively weak attention at the back of the focal.

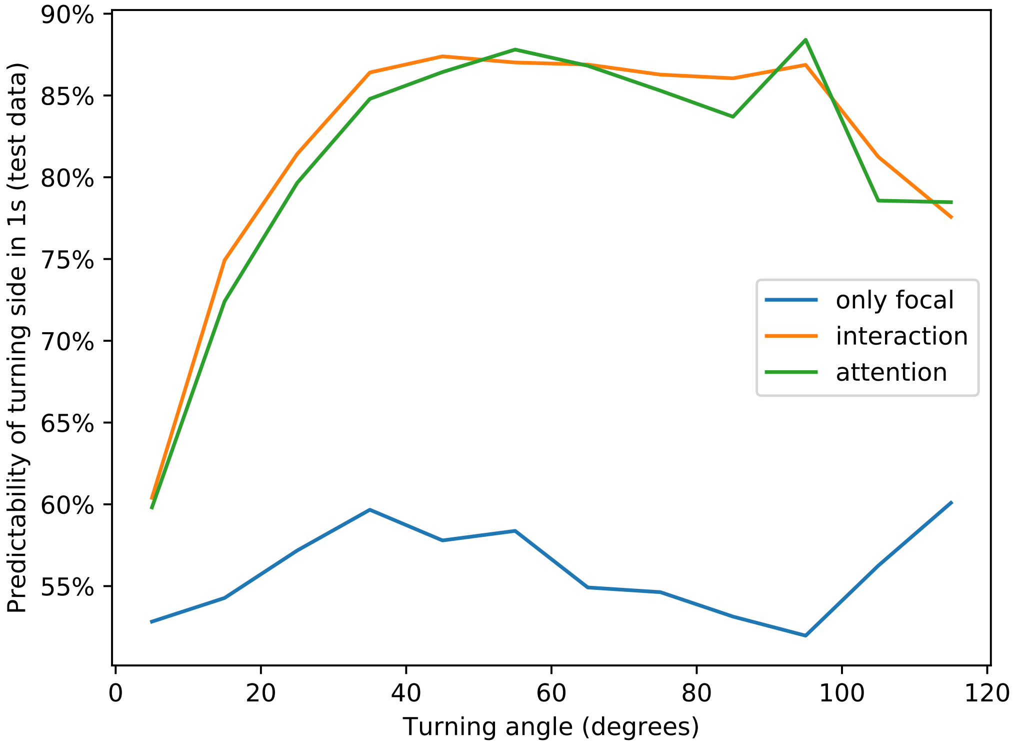

Figure S9: **Accuracy in the prediction of which side a fish is turning towards in 1s as a function of angle turned by the fish after 1s.** Accuracy of a network using only focal variables (blue) remains low at all turning angles. Both networks integrating information from 25 neighbours, interaction (orange) and attention (green) perform better at turning angles between 40 and 100.

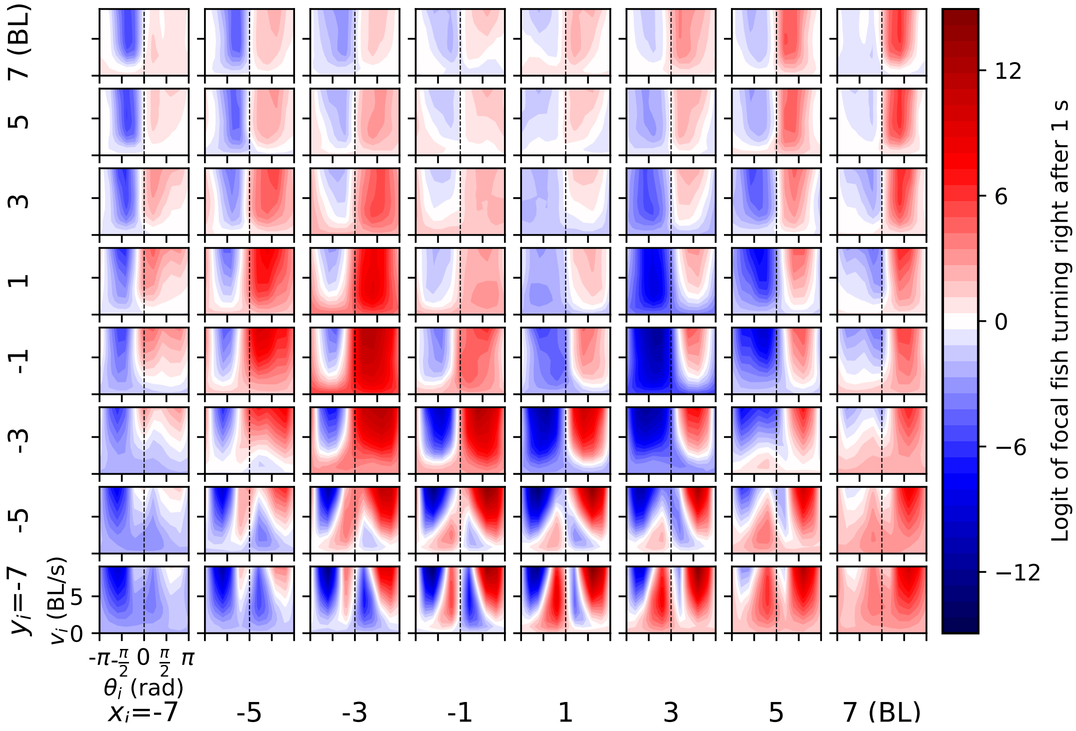

Figure S10: **Properties of interaction between a pair of fish in the collective when the focal fish is moving at low speed**. As Figure 2B in main text but for focal speed fixed to 1 BL/s.

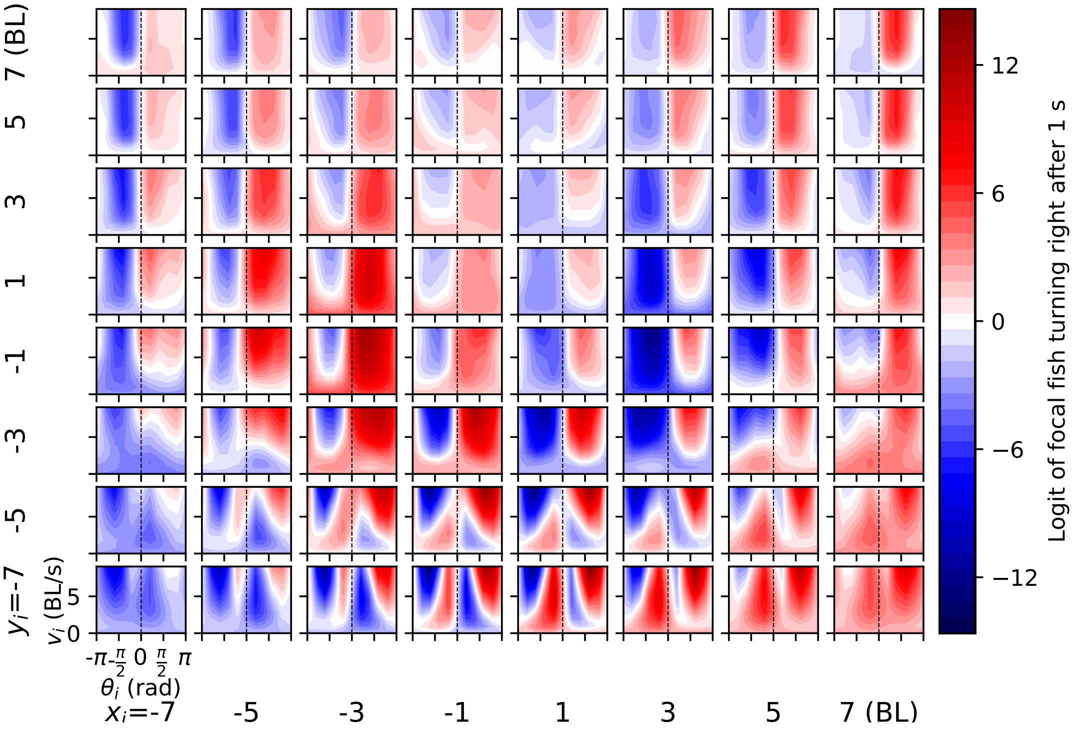

Figure S11: **Properties of interaction between a pair of fish in the collective when the focal fish is moving at medium-low speed**. Same as Figure S2A but focal speed is fixed to $v = 2$ BL/s.

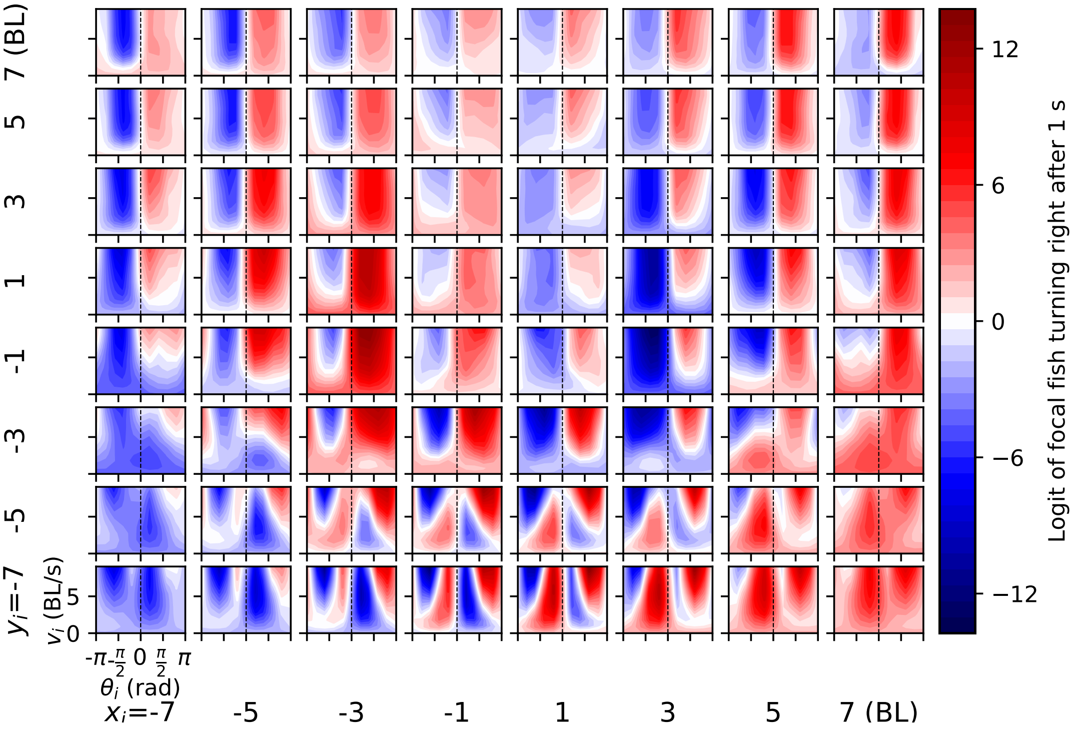

Figure S12: **Properties of interaction between a pair of fish in the collective when the focal fish is moving at medium-high speed**. Same as Figure S2A but focal speed is fixed to $v = 4$ BL/s.

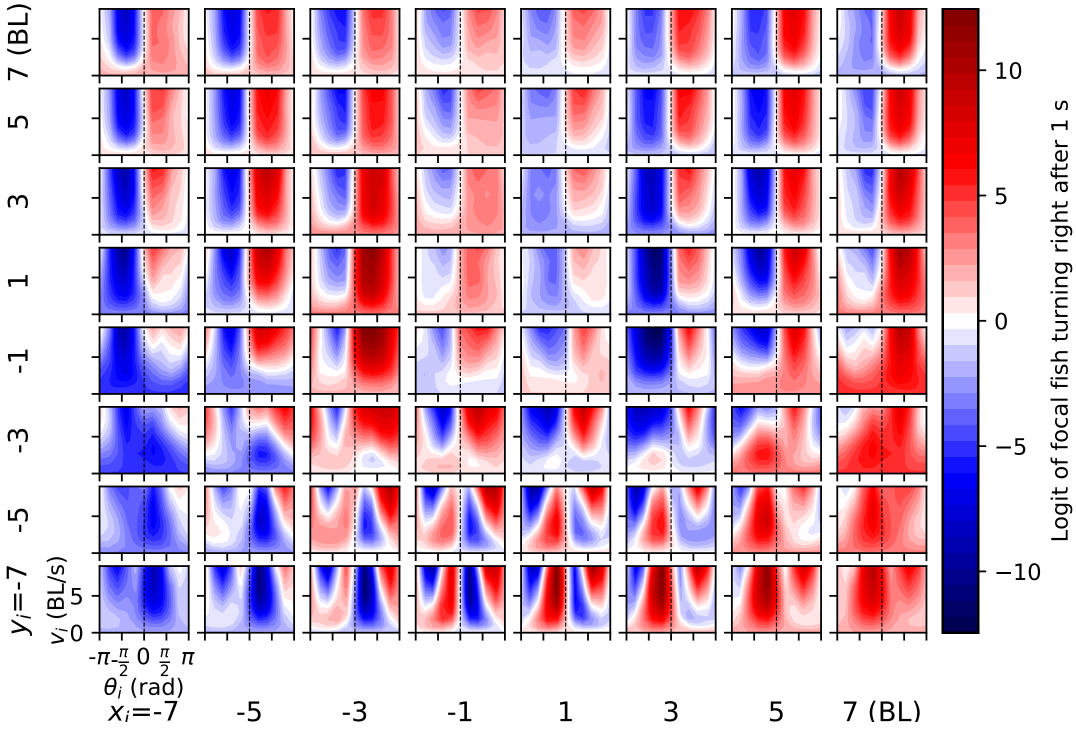

Figure S13: **Properties of interaction between a pair of fish in the collective when the focal fish is moving at high speed**. Same as Figure S2A but focal speed is fixed to $v = 8$ BL/s.

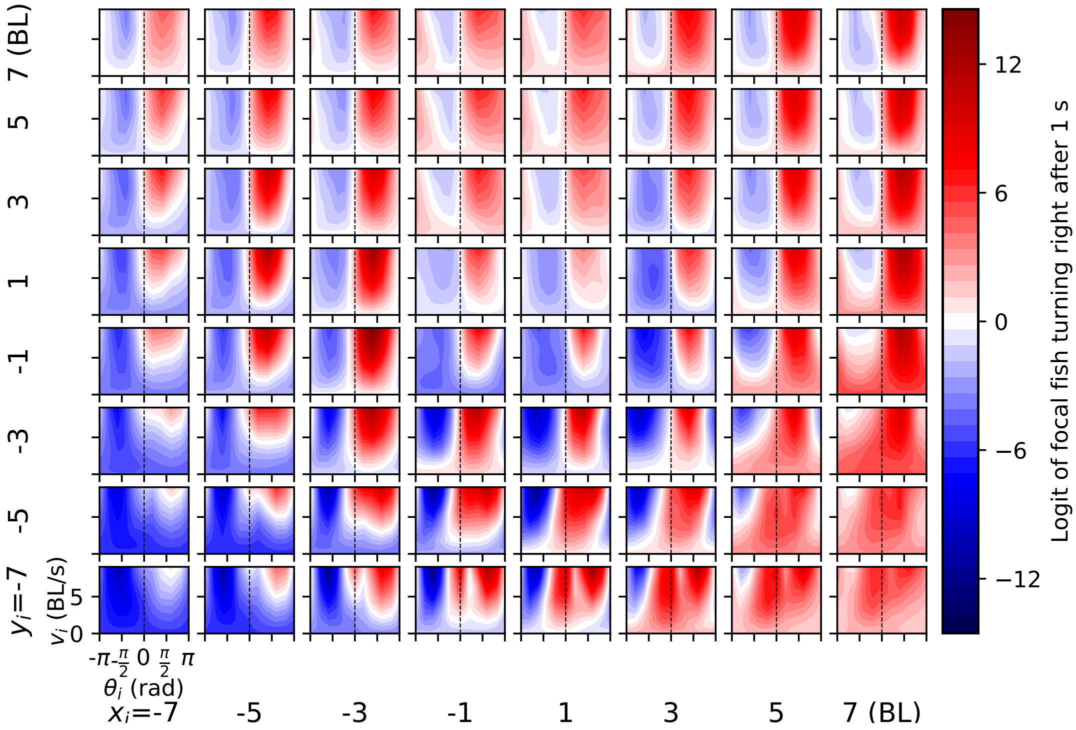

Figure S14: **Properties of interaction between a pair of fish in the collective when the focal fish is in the midst of a right turn**. Same as Figure 2B in main text but with focal normal acceleration fixed to $a_\perp = 100$ BL/s^2.

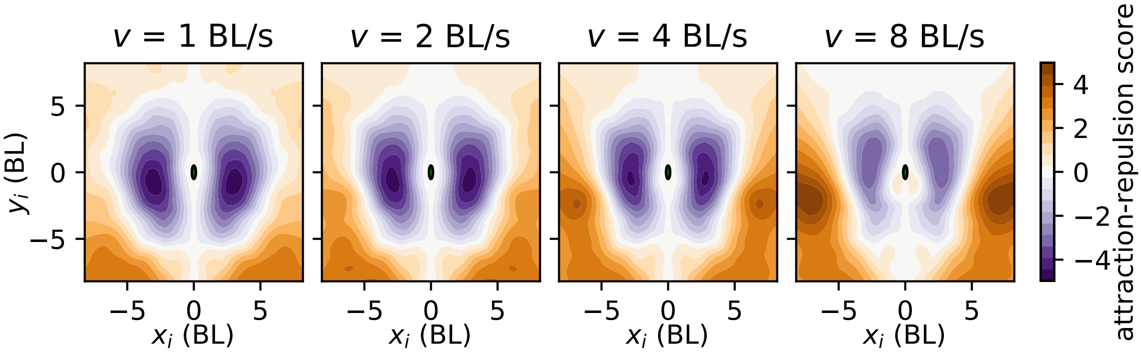

Figure S15: **Relative attraction and repulsion zones depend on kinematic parameters of focal** Like Fig. 3C but at four different focal speeds (1, 2, 4 and 8 BL/s) while keeping neighbour speed fixed at the median speed, v_i = 3.04 BL/s

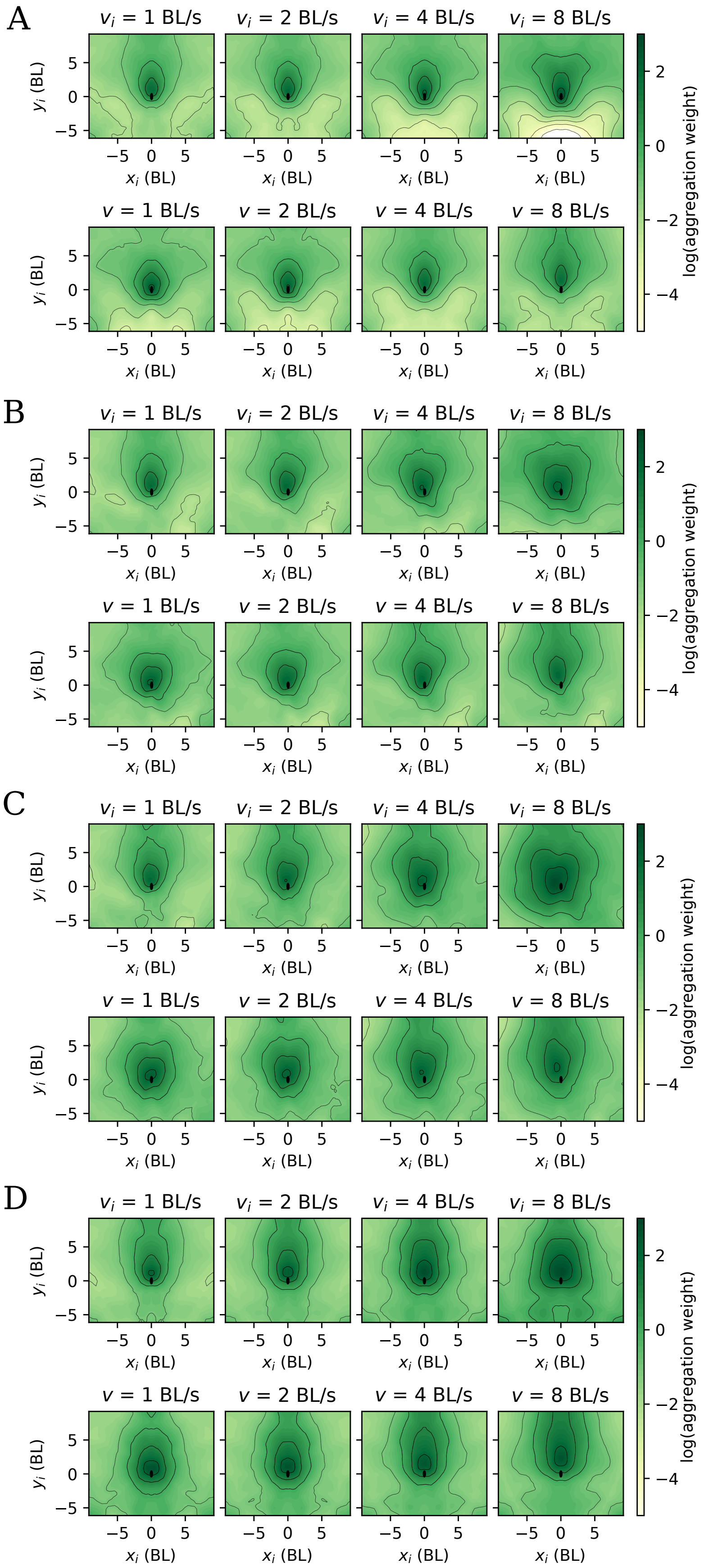

Figure S16: **Aggregation with information about relative orientation**. Same as Fig. 4A, but when the attention subnetwork is trained with the relative orientation of the neighbour, in addition to the variables used in the main text. **A** The neighbour is parallel (at 0 degrees) to the focal. **B** The neighbour is at 45 degrees (towards the right) with the focal, **C** the neighbour is perpendicular (90 degrees) and pointing to the right of the focal. **D** The neighbour is antiparallel (180 degrees) to the focal.