Science Advances
AAAS

Supplementary Materials for

Subcritical escape waves in schooling fish

Winnie Poel *et al.*

Corresponding author: Pawel Romanczuk, pawel.romanczuk@hu-berlin.de

Sci. Adv. **8**, eabm6385 (2022)
DOI: 10.1126/sciadv.abm6385

This PDF file includes:

1 Construction of rescaled interaction networks

This section describes how the interaction networks for densities that were not observed in experiments were constructed. Briefly said, we

- first rescaled the position data,
- then approximated each individual fish by an ellipse,
- used an active particle simulation to ensure that when placed at the rescaled positions using the original orientations, the ellipses would relax into a non-overlapping configuration,
- then calculated the visual field of each ellipse using an analytically derived formula and afterwards accounting for occlusions using an algorithm based on the analytic results to determine the angle which one ellipse occupies in the visual field of another ellipse,
- before finally generating the interaction network based on the calculated values of log metric distance and ranked angular area.

We conclude this section with a comparison of the properties of the thus constructed networks with the networks previous studies [27,28] constructed based on bodyshape reconstruction from tracking software and raycasting. A comparison of the two types of networks based on their ability to describe the observed cascade size distributions when used together with the behavioral contagion process can be found in the next section.

1.1 Rescaling position data

The position data was rescaled with the following range of factors: $\lambda \in [0.3, 0.4, \ldots, 2.8, 2.9]$ setting

$$\begin{aligned} x_{\text{rescaled}} &= \lambda \cdot x_{\text{exp}} \\ y_{\text{rescaled}} &= \lambda \cdot y_{\text{exp}} \ . \end{aligned} \tag{S1}$$

This yields positions with a range of median nearest distances. Figure S1 depicts histograms of the number of networks per interval of median nearest neighbor distance (of width 0.11 BL, as used in Figure 2b of the main text) for both the 'Baseline' and the 'Alarmed' dataset as well as the larger groups from [27] (see section 3.8).

1.2 Ellipse approximation

To approximate a single fish we used an ellipse with its center at (x_0, y_0), semi-major axis of length a and semi-minor axis of length b. The semi-major

axis is rotated by the fish's orientation, ϕ, from the x-axis. We call $w = b/a$ the aspect ratio of the ellipse. The ellipse has a single eye that is positioned on the semi-major axis at a distance l from the center of the ellipse, where $l = 1$ refers to the front, $l = 0$ the center and $l = -1$ to the back of the ellipse (see Fig. S2). We chose $w = 0.14$ and $l = 0.9$ for all fish in agreement with the average value obtained from tracking software data where we used the ratio of detected inter eye distance and bodylength of a fish as w and the position of the eye along the detected midline to determine l, see Fig. S3. Because the tracking software saved the edge of the head of the fish as its position $(x_{\text{exp}}, y_{\text{exp}})$ we positioned the ellipse at

$$\begin{aligned} x_0 &= x_{\text{exp}} - \frac{a}{2}\cos\phi \ , \\ y_0 &= y_{\text{exp}} - \frac{a}{2}\sin\phi \ . \end{aligned} \tag{S2}$$

1.3 Active particle simulation to eliminate overlaps

Since we assume our school of fish to be 2 dimensional there is an upper limit to the density we can achieve by rescaling (see also section 3.2 of the SI). We acknowledge this by ensuring that no two fish overlap in the rescaled position data using an active particle simulation based on the code provided within [71] Ellipses are placed at the rescaled positions at $t = 0$ with velocities $v_i(t = 0) = 0$. Their velocities $\mathbf{v}_i(t)$ change based on their positions $\mathbf{x}_i(t) = (x_i(t), y_i(t))^T$ according to

$$\begin{aligned} \frac{d\mathbf{v}_i(t)}{dt} &= \frac{1}{\mu_i}\left(-\alpha\mathbf{v}_i(t) + \sum_{j\neq i}^{j} \frac{\mathbf{x}_i - \mathbf{x}_j}{|\mathbf{x}_i(t) - \mathbf{x}_j(t)|} F_{ij}(t)\right) \\ F_{ij} &= \lambda A_{ij}(t) \end{aligned} \tag{S3}$$

where $A_{ij}(t)$ is the overlap area of the ellipse i and ellipse j at time t, $\mu_i = \mu = 1$ is the mass of the ellipse, α the damping parameter and λ a constant model parameter. We use $\alpha = 0.2$ and $\lambda = 0.05$ and stop the simulation as soon as no more overlaps are detected. To speed up the relaxation into a non-overlapping state, we used a second larger ellipse (factor 1.1) to determine the repulsion area A_{ij} and stopped when the original size ellipses were no longer overlapping, see Fig. S4.

1.4 Analytical calculation of visual field of ellipse

In order to determine the interaction network of the ellipses we need a matrix of ranked angular areas and distances for all pairs of ellipses (see eq. S11). The ranked angular areas were determined based on the analytical calculation of the angular area that an ellipse at a relative position (x_0, y_0), semi-major axis of length a rotated by ϕ from the x-axis and semi-minor axis of length b has for an observer sitting at the origin. We assume the observer has 360° vision. The observed ellipse is given by:

$$\begin{aligned} \begin{pmatrix} x \\ y \end{pmatrix} &= \begin{pmatrix} x_0 + a\cos\psi\cos\phi + b\sin\psi\sin\phi \\ y_0 + a\cos\psi\sin\phi - b\sin\psi\cos\phi \end{pmatrix} \\ &\text{with} \quad 0 \leq \psi < 2\pi \ . \end{aligned} \tag{S4}$$

To describe the fish we use the aspect ratio w and polar coordinates r, θ

$$\begin{aligned} a &= 1/2 \\ b &= w/2 \\ x_0 &= r\cos\theta \\ y_0 &= r\sin\theta \ . \end{aligned} \tag{S5}$$

We can then calculate the gradient of the ellipse as

$$\frac{dy}{dx} = \frac{\sin\phi\sin\psi + w\cos\phi\cos\psi}{\cos\phi\sin\psi - w\sin\phi\cos\psi} \ . \tag{S6}$$

A tangent line to the ellipse is given by

$$y = \left.\frac{dy}{dx}\right|_s (x - x_s) + y_s \tag{S7}$$

where (x_s, y_s) is the tangent point on the ellipse (see Fig. S2).

Since we place our observer at the origin, the tangents we need in order to determine the visual field have to pass through this point and thus it needs to hold that

$$\begin{aligned} 0 &= \left.\frac{dy}{dx}\right|_s (0 - x_s) + y_s \\ 0 &= -\left.\frac{dy/d\psi}{dx/d\psi}\right|_s x_s + y_s \\ 0 &= -(dy/d\psi)|_s x_s + (dx/d\psi)|_s y_s \end{aligned} \tag{S8}$$

which can be written as

$$0 = \begin{vmatrix} x_s & dx/d\psi|_s \\ y_s & dy/d\psi|_s \end{vmatrix}$$

$$\begin{aligned} x_s &= r\cos\theta + \frac{\cos\phi\cos\psi + w\sin\phi\sin\psi}{2} \\ \frac{dx}{d\psi}|_s &= \frac{-\cos\phi\sin\psi + w\sin\phi\cos\psi}{2} \\ y_s &= r\sin\theta + \frac{\sin\phi\cos\psi - w\cos\phi\sin\psi}{2} \\ \frac{dy}{d\psi}|_s &= -\frac{\sin\phi\sin\psi - w\cos\phi\cos\psi}{2} \end{aligned} \tag{S9}$$

where $||$ is the determinant. Solving for ψ yields

$$\begin{aligned} \psi_\pm &= \pm\, 2\tan^{-1}\left(\frac{\gamma \mp r\sin(\theta-\phi)}{\beta}\right) \\ \beta &= w(2r\cos(\theta-\phi) - 1) \\ \gamma &= \big(-w^2 + 2r^2\left((1+w^2)\right. \\ &\quad \left.+(w^2-1)\cos(2(\theta-\phi))\right)\big)^{1/2} \ . \end{aligned} \tag{S10}$$

Inserting (S10) into (S4) returns the tangent points (x_{x1}, y_{s1}) and (x_{x2}, y_{s2}) whose polar angles, θ_{s1} and θ_{s1}, determine the angular area of the ellipse, $\alpha = \min(|\theta_{1s} - \theta_{2s} + n\pi|, n \in \mathbb{N})$ (here we use the fact that the angular area needs to be smaller than π because the ellipses aren't allowed to overlap).

Occlusions of individual j in the visual field of i by all other individuals $k \neq j$ are then determined by an algorithm using intersections of rays originating from the eye of the focal individual i and going through the tangent points (as perceived by i) on j with the outlines of ellipses k. To determine if j or k is visible to i in the occluded area, intersection of ellipse outlines with the angle bisectors of the rays described above are also considered.

1.5 Network construction

As described in the methods section of the main paper, the weight of a link (from i to j) is given by:

$$w_{ij} = (1 + \exp(-\beta_1 - \beta_2 LMD - \beta_3 RAA))^{-1} \ . \tag{S11}$$

LMD is the log of the metric distance between ellipse i and j and RAA is the ranked angular area of j in the visual field of ellipse i. Coefficients β_i were obtained from a logistic regression of experimental observations of first responders and can be found in table S2. Data from before and after exposure to Schreckstoff were fitted with one set of coefficients in accordance with [28] who did not find any significant change from 'Baseline' to 'Alarmed' state.

1.6 Density dependence of network measures

Fig. S5 shows different network properties plotted against median nearest neighbor distance (NND) in body lengths. For more details refer to the figure caption.

2 Model Calibration

As described in the methods section of the main paper, the behavioral contagion model has one free parameter, the response threshold, that corresponds to an individual responsiveness to social cues. We calibrate the model separately for each dataset using the experimentally observed cascade sizes following a maximum likelihood approach as in [28]. This allows us to determine which value of the parameter response threshold best describes the observed data.

Each startle cascade observed in the experiments corresponds to one interaction network based on the positions and visual fields of all individuals at the time just before the initial startle. In our model we artificially initiate a startling cascade on this network by setting the node corresponding to the experimentally observed initial startler to the active state at $t = 0$ and recording the resulting cascade size (i.e. the number of individuals, including the initial startler, changing their state to active before the cascade dies out and no active individuals remain). This is done 10000 times for each interaction network and for a range of response threshold values to build a distribution of cascade sizes for each response threshold and network. These probability distributions combined with the corresponding experimentally observed cascade size allow us to determine which response threshold is most likely given our observed data (i.e. has the maximum likelihood). The result can be seen in Fig. S6 for both

the networks constructed using ellipses to approximate body shape and visual field (solid lines) as well as networks based on the tracking software described in [27,28].

Fig. S7 shows the calibrated model and the experimental data it describes.

3 Model assumptions and limitations

3.1 Assessing the quality of ellipse based networks

To see the effects of our approximation of bodyshape and field of view using ellipses and analytical calculation we first compare the resulting networks via their link weight distribution, as well as their (weighted and binary) degree distribution. Then, to see the influence our approximation may have on the dynamics of the complex contagion process and the model's ability to describe the observed data, we fit the single free parameter of the model, the response threshold, once for the (original scale) ellipse networks and once for the networks constructed using the ranked angular area determined through ray casting.

The distributions of network properties for both types of networks are shown in Fig. S8 where the ellipse networks are drawn as a line and the ray casting networks as a shaded area. The histograms show (from left to right): weighted degree, binary degree, link weights, link distance (metric distance between two fish connected by a link). Each row corresponds to one of the datasets, top: 40 fish 'Baseline', middle: 40 fish 'Alarmed', bottom: 150 fish. Overall, the distributions are generally in good agreement. The remaining differences between them are due to the fact, that for the ellipse networks all overlaps of individuals were eliminated to ensure a completely 2 dimensional school whereas in the experiments the shallow water still allowed the fish to occasionally cross path and appear stacked on top of each other in the data. The elimination of these overlaps by shifting the individuals apart leads to on average larger inter-individual distances in the ellipse networks, as seen in the right column of Fig. S8. Additionally, putting these originally stacked individuals next to each other in 2 dimension now means that they occlude each other's field of view, on average leading to fewer visible neighbors and thus fewer links per individual. This results in a shift in the weighted and binary degree distributions (two left columns, fig. S8) while the distribution of link weights does not change (second right column). The effect is especially prominent for the high density case (40 fish 'Alarmed', second row) in which overlaps where most frequent.

The results of the maximum likelihood fitting of the response threshold for both network constructions can be found in Fig. S6. A plot of the experimental data with the cascade size distribution produced by the fitted models is shown in Fig. S7.

3.2 Density limit

The regulation of group density has a naturally occurring limit towards high density because of the physical bodies of the fish. We need to consider this as a limit to the group's ability to modulate responsiveness solely via group density. When taken to the extreme this limit becomes the problem of closely packing ellipses of aspect ratio 0.14, but this limit certainly is not of biological relevance, especially in our model which is based on visual interactions. Instead we take a look at all the configurations we find in our original data (no rescaling) after ensuring that all overlaps have been resolved via the active particle simulation. The lower limit of density (measured by median nearest neighbor distance of the group) occurring in our experiments gives us an indication to what densities are possible as a group median, namely (40 fish: 0.44, 150 fish: 0.48), while individual fish can get closer (40 fish: 0.16, 150 fish: 0.15). Additionally, we constructed the density dependence of average number of physical contacts per individual from all the rescaled and original data we have (Fig. S9). When a large fraction of the ellipses start touching the limit of our vision based model is most likely reached as interaction can no longer be viewed as purely visual but become governed by physical forces and other sensory inputs.

3.3 Two dimensional schools

As in [27] and [28] we assume schools to be approximately 2 dimensional. As pointed out in the previous section, this yields a lower limit for the nearest neighbor distance (NND) of a school and

by this an upper limit for the branching ratio that can be reached without an additional change in individual response threshold. In the main text and Fig. S14 we find that this density limitation does not allow the schools to cross the critical manifold by a change of NND alone. Here, we make a simple argument to explain why we expect the maximum branching ratio that could be reached in a 3 dimensional school via a change of NND to be similar to that in 2 dimensions.

As described in the methods section of the main paper, the branching ratio is calculated as

$$b_j = \frac{\tau_{\text{act}}}{\theta_{\max}} \sum_i \frac{w_{ij}}{K_i}. \tag{S12}$$

The branching ratio of an individual is thus maximized when the ratio of the weight of its incoming networks links, $w_j = \sum_i w_{ij}$ and the absolute number of its incoming network links K_j is maximized. This is the case, when the average strength of a link in the network is largest, which is the case for small NND because of the functional dependence of w_{ij} on interindividual distance.

Thus because of the fractional contagion process and the resulting K_i in the denominator of equation (S12), it is the average link strength determining the maximal branching ratio at high densities and not the total weight of incoming networks links. For a densely packed configuration of ellipses, we expect the average link strength (corresponding to the links to the nearest neighbors) to be similar in 2 or 3 dimensions and thus the average branching ratio should be similar. A more detailed analysis would require the reconstruction of visual fields in 3 dimensions and is beyond the scope of this paper.

3.4 Visual threshold

Because of limitations of the visual sensory perception and cognitive limitations it seems reasonable to assume that the angular area an individual i occupies in the visual field of individual j, α_{ij} must be larger than a certain value, $\alpha_{ij} > \alpha_{min}$ in order for individual i to be seen by and influence the behavior of individual j. In the main text we use $\alpha_{min} = 0.02$. Here, we explore further values. We construct networks and fit the individual response threshold of the model for each threshold value. Table S3 gives these fitted individual threshold parameters. When larger visual thresholds are used, the average response threshold best describing the experimental data increases. The increased visual threshold lowers the average degree of a node, K_i, and thus because received cues are weighted with $1/K_i$ (see methods section of main text), received cue intensity becomes larger on average. To compensate for this effect, the average response threshold must be increased. Fig. S10 shows the dependence of the average branching ratio for the different visual thresholds. At high densities networks are similar for all thresholds because basically all neighbors are very close by and occupy a large angular area in the visual field and smaller (further away) individuals are blocked from view. The increased average response thresholds for larger visual thresholds then lead to a smaller branching ratio at high densities. Our main finding that observed schools are subcritical is not changed by this and neither is the general form of the relative individual payoff.

3.5 Choosing initiators

As seen in Fig. S11 our results do not change qualitatively when choosing casacade initiators as network neighbors instead of randomly as in the main text.

3.6 Choice of agent memory

For our study we have differed from [28] in our choice of agent memory by choosing $\tau_m = 1$ instead of $\tau_m = 2$. Fig. S12 shows that our results do not critically depend on this choice.

3.7 Influence of polarization

It has to be acknowledged that while median nearest neighbor distance appears to be capturing the most relevant feature of the spatial structure of the group, there are of course other measures that could be considered. A commonly used measure in collective behavior is the group polarization (degree of orientation order), defined as the normalized absolute sum of the orientation unit vectors of all individuals, $\zeta_i = (\cos(\phi_i), \sin(\phi_i))^T$, as

$$\Phi = \frac{1}{N} \left| \sum_{i=1}^{N} \zeta_i \right| . \tag{S13}$$

It can take values between 0 and 1, where 1 represents perfect alignment of individuals and large systems of individuals with completely uncorrelated orientations tend towards a polarization of 0. Figure S13 shows the polarization of the schools at the beginning of each observed startle cascade and the corresponding cascade size. While the data sets are not large enough for a quantitative analysis, there could be positive correlation between polarization and cascade size. Especially, a high polarization could explain the observed deviation of the largest two cascades in the dataset of larger group size (150 fish) from the model prediction (see Figure S7).

3.8 Larger group size

The dataset used in [27] of groups of approximately 150 fish shows a distribution of median nearest neighbor distances which resembles that of the 'Alarmed' state of the smaller groups of 40 individuals (see Figure S1) and contains $n = 134$ observed startle cascades. Figures S6 and S7 show the observed relative frequency of cascade sizes and the fitted model. The fitted value of the average response threshold for this dataset is significantly smaller than those of the other two datasets. The model may visually appear to underestimate large cascades (see the deviation of the two largest observed cascade sizes). However, when accounting for the variation of model predictions for the different networks (shaded green area) and the sampling error of the small number of observed cascades (error bars on experimentally observed relative frequencies of cascade sizes), the deviation appears less drastic. In addition, the two data points show a high polarization value (see Figure S13) which may introduce some aspects into the collective behavior that are not adequately captured by the model due to a lack of data to systematically analyze this effect.

The coefficients of the logistic regression for the first responder probability are taken from the original study and given in table S2. They deviate from the ones for the smaller data set as well.

Fig. S15 shows the branching ratio for this dataset. The majority of the observed schools (green dots in subplot A) are subcritical. However, some of the networks with the lowest median NND show a supercritical branching ratio, $b > 1$. Even though the smallest median NND values measured in the larger groups are comparable to that of the 'Alarmed' dataset, the lower average response threshold yields a higher branching ratio at comparable NNDs (see equation (6) in the main text, where $b \propto \theta_{\max}^{-1}$). In addition, the deviation in the logistic regression coefficients, which enter the calculation of link weights from spatial positions, affect the dependence of the average branching ratio on the median NND. Overall it seems that the larger groups are overall closer to the critical point than the smaller groups.

It is however important to note that the two datasets (small/large groups), while using the same species, differ not only in group size (40 vs. 150 ± 4 fish), but also in experimental procedure and setup (i.e. handling of the fish prior to being placed in the tank, recording time, size- and age-matching, size of the tank and tank area per fish). All of these factors could cause the difference in average individual response threshold, the observed median nearest neighbor distances and coefficients of the logistic regression (see Table S2) between the datasets. It is therefore not possible to attribute the decreased distance to criticality to the increased group size. However, the hypothesis that increased group size decreases distance to criticality should certainly be addressed in future studies.

3.9 Discussion of model limitations and further experimental studies

There are a number of limitations of the behavioral contagion model that could be addressed by more experimental data to improve and refine statistical analysis. Currently, interaction networks are built on pair-wise interactions fitted to first responder data. This assumes that interactions do not change during the cascade and that higher order interactions do not play a significant role. Additionally, with more experimental observations we could determine the features predictive of startle response and their relative importance for data confined to small density ranges and thus test if interaction rules depend on density. This would potentially show other types of interactions (acoustic, sensomotor) playing a role at higher densities. This detailed data for different (naturally occurring) densities could also provide a direct observation of the increase in average cascade size with increasing den-

sity. In an experimental setup that allows us to startle 1 or 2 initial startlers a detailed analysis of resulting cascade size distribution could potentially even show the increase in sensitivity with density as in Figure 3B and D of the main text or Fig. S11.

The relative payoff measure is simplistic in the sense that it focuses solely on information processing which is only a single aspect of a wider range of important factors and lacks experimental evidence for the assumed reaction to a predator. Experiments with a real predator are needed to determine the number of initial startlers and potentially even see the density dependence of predator detection. Providing a stimulus in empty tank at different densities that occur via natural fluctuations may present another possibility test if density indeed has an effect on the detection ability. This lends itself to investigate another aspect: In our model we assume that the detection of the predator happens exclusively at the beginning of the cascade and the information then spreads only socially. In reality there will always be a direct detection of the predator happening along side the behavioral contagion process that reinforces the spreading and might be the reason that we observe such high false negatives in our model, which does not have this mechanism. In experiments with an artificial stimulus one could potentially change the duration for which a stimulus is shown to quantify this effect.

4 Sensitivity and branching ratio

4.1 Variation within original scale data

Figure S16 shows the model predictions for the original scale networks for both measures of criticality, the collective sensitivity and the branching ratio, as scattered data points (each point represents one network which corresponds to an experimental observation). One observes that due to the variance in observed spatial configurations some of the networks in the 'Alarmed' dataset can be considered critical when using the branching ratio as an estimate while their collective sensitivity remains below the maximum.

4.2 Alternative definitions of collective sensitivity

Fig. S11 depicts differences between average cascade sizes initiated by m and $m+1$ or 1 and m initial startlers as a generalization of the definition in the main text. These sensitivities, while not directly motivated by the underlying theory of criticality, may be biologically relevant. We find, that independent of the used definition, we observe a maximum in collective sensitivity close to criticality. We note that the schools are most sensitive to differences for small numbers of initial startles.

4.3 Branching ratio

The dependence of the branching ratio on median nearest neighbor distance and the maximum response threshold is shown in Fig. S14 and S15 as well as the averages over original scale networks, characterizing the experimentally observed schools as on average subcritical.

4.4 Estimating the branching ratio from first responder times

While in our manuscript we base the calculation of the branching ratio on the inferred interaction networks and knowledge about the contagion process, a different approach is to try and directly measure it from the observed cascades. Here, the quantity one aims to measure is the average number of fish that startle in response to a single startle. This raises the need to assign causality between startles, e.g. by defining a typical time interval and assuming that all startles in the current interval were caused by startles in the last interval. Without the detailed model including the reconstructed interaction networks, it is however not easily possible to assign causality for startle cascades, because response times of fish to an observed startle vary largely and because any startle after the first responder (i.e. the second startle) is potentially influenced by all other startles before due to the high speed of the cascade propagation. Figure S17A depicts the first responder times, i.e. the time between the first and the second startle in a cascade, where causality is most clear. It is not clear how to define a typically time step based on the wide distribution of response times. Figure S17B therefore depicts the number of

fish startled within a time interval Δt after the initial startle averaged over all observed cascades in each of the datasets as a function of the time interval Δt (solid lines). The estimate strongly depends on the choice of the time interval and therefore can only be estimated with large uncertainty. However, for a choice of Δt larger than 50 to 75% of the observed first responder times ($0.1\,\mathrm{s} \leq \Delta \leq 0.2\,st$ as indicated by vertical lines), both estimates roughly agree.

4.5 Cascade size distributions near criticality

Figure S18 depicts cascade size distribution for the combined networks of 40 individual as well as the larger group sizes for a maximum response threshold of $\theta_{\max} = 0.016$ (comparable to that of the 150 fish data) near the critical point. Each distribution is the result of combining the simulations runs (1000 per network) of all networks within a bin of median NND (bin width approximately 0.05 BL). Bins typically contain about 200 networks for the small group size and 50 for the larger groups. For the yellow distribution the bin is centered around the critical median NND resulting in $b = 1$, while the dark blue distribution is centered around the median NND that yields the maximal collective (1v2) sensitivity. Using the average branching ratio, a critical median NND of 0.74 BL for the 40 fish and 0.65 BL for the 150 fish data is estimated. Using the maximum in sensitivity, the estimated critical median NND are 0.50 and 0.45 for the 40 and 150 fish data respectively. The distributions show that the estimated critical median NNDs give a good upper and lower bound of the critical point, which is indicated by a power law distribution of relative cascade sizes.

5 Relative payoff measure

We find in the main text that the relative payoff shows two local maxima. One is a result of the maximal visual detection of the predator by group members (as seen in Fig. S20 where that maximum increases with increasing the fraction of individuals responding to their personal visual predator detection by startling). The other is due to the peak in sensitivity (as seen in the main text or Fig. S11) to the number of initial startlers at the critical point (see Fig. S21, where the initial response to predator and noise cue is assumed to be identical and there is no maximum at criticality). The position of the maximum payoff depends on the relative noise cost but also on the parameters of the visual predator detection as we will explain here before giving a more detailed derivation of the payoff measure.

5.1 Parameters of visual predator detection

Our model of visual predator detection depends on three parameters, namely the distance of the predator from the group boundary, d_{pred}, the maximal distance at which a predator is still visible $d_{\max}$ and the response probability of an individual p_{detect} (probability to startle given that the individual has visual access to the predator). Figures S20 and S24 show the influence of p_{detect} on the relative payoff. Fig. S22 illustrates how $d_{\max}$ and d_{pred} influence the number of individuals that can see the predator, Fig. S23 how they influence the relative payoff. In Fig. S23A one does not observed curves with two separate maxima but just one maximum shifting from a density optimizing individual access to visual information of the predator to the critical density, optimizing sensitivity to number of initial startlers. The observed median NNDs of the 'Baseline' and the 'Alarmed' dataset are optimal for a relative noise cost of $\xi \approx 3$ and $\xi \approx 1$ respectively.

5.2 Construction of the relative payoff measure

Here, we start by considering all possible behavior-environment combinations (false and true positives and false and true negatives, marked fp, tp, fn and tn respectively), their costs κ and the rates with which they occur ρ. All of them are added up into a payoff rate defined as

$$\tilde{\psi} = \rho_{\mathrm{fp}}\ \kappa_{\mathrm{fp}} + \rho_{\mathrm{fn}}\ \kappa_{\mathrm{fn}} + \rho_{\mathrm{tp}}\ \kappa_{\mathrm{tp}} + \rho_{\mathrm{tn}}\ \kappa_{\mathrm{tn}} \quad \text{(S14)}$$

We assume that a predator causes $N_{\mathrm{init}} = n(\mathrm{NND})$ initial startlers while a noise cue causes just one, $N_{\mathrm{init}} = 1$, and cues indicating predators appear at a rate ρ_{pred} while cues indicating no predator appear at rate ρ_{noise}. Then, each individual has the

following rates of

$$\begin{aligned}
&\text{false pos.:} \quad \rho_{\text{fp}} = \rho_{\text{noise}}\; p(\text{startle}|N_{\text{init}} = 1),\\
&\text{false neg.:} \quad \rho_{\text{fn}} = \rho_{\text{pred}} \left[1 - p\big(\text{startle}|N_{\text{init}} = n(\text{NND})\big)\right],\\
&\text{true pos.:} \quad \rho_{\text{tp}} = \rho_{\text{pred}} - \rho_{\text{fn}},\\
&\text{true neg.:} \quad \rho_{\text{tn}} = \rho_{\text{noise}} - \rho_{\text{fp}},
\end{aligned} \tag{S15}$$

where $p(\text{startle}|N_{\text{init}})$ means the probability to startle (as part of the cascade) given N_{init} initial startlers and can be obtained from simulations. Inserting (S15) into (S14) yields

$$\begin{aligned}
\tilde{\psi} =& \rho_{\text{pred}}\kappa_{\text{tp}} + \rho_{\text{noise}}\kappa_{\text{tn}} + \\
&\rho_{\text{fp}}(\kappa_{\text{fp}} - \kappa_{\text{tn}}) + \rho_{\text{fn}}(\kappa_{\text{fn}} - \kappa_{\text{tp}}).
\end{aligned} \tag{S16}$$

Since we are interested in the relative payoff of different school densities for a fixed environment and do not want to compare payoff rates between different environments, we can choose the baseline freely. We set it to $\rho_{\text{pred}}\kappa_{\text{tp}} + \rho_{\text{noise}}\kappa_{\text{tn}}$, the rate at which an individual on average gains benefits by making correct decisions. By rescaling the payoff rate in units of $\rho_{\text{pred}}(\kappa_{\text{fn}} - \kappa_{\text{tp}})$ (the average payoff rate associated to a predator cue) we can define a relative payoff as

$$\begin{aligned}
\psi &= \frac{\tilde{\psi} - (\rho_{\text{pred}}\kappa_{\text{tp}} + \rho_{\text{noise}}\kappa_{\text{tn}})}{\rho_{\text{pred}}(\kappa_{\text{fn}} - \kappa_{\text{tp}})} \\
&= \frac{\rho_{\text{fp}}}{\rho_{\text{pred}}} \left(\frac{\kappa_{\text{fp}} - \kappa_{\text{tn}}}{\kappa_{\text{fn}} - \kappa_{\text{tp}}}\right) + \frac{\rho_{\text{fn}}}{\rho_{\text{pred}}}
\end{aligned} \tag{S17}$$

Inserting S15 then yields

$$\begin{aligned}
\psi = \xi\; & p(\text{startle}|N_{\text{init}} = 1) \\
&+ [1 - p(\text{startle}|N_{\text{init}} = n(\text{NND}))]
\end{aligned} \tag{S18}$$

with

$$\xi = \frac{\rho_{\text{noise}}}{\rho_{\text{pred}}} \left(\frac{\kappa_{\text{fp}} - \kappa_{\text{tn}}}{\kappa_{\text{fn}} - \kappa_{\text{tp}}}\right) \tag{S19}$$

Here, $\rho_{\text{noise}}/\rho_{\text{pred}}$ is the relative prevalence of noise cues compared to the predator cues. If $\rho_{\text{noise}}/\rho_{\text{pred}} \gg 1$ the environment is very noisy, if $\rho_{\text{noise}}/\rho_{\text{pred}} \ll 1$ there is a lot of predation. In the cost-based term $\frac{\kappa_{\text{fp}} - \kappa_{\text{tn}}}{\kappa_{\text{fn}} - \kappa_{\text{tp}}}$ the numerator measures the costs associated to startling behavior, the denominator quantifies the costs of an attack that are due to injury or risk of death. This term can be thought of as the relative costs of noise. The combination of both terms, ξ, can best be described as the *relative noise cost*.

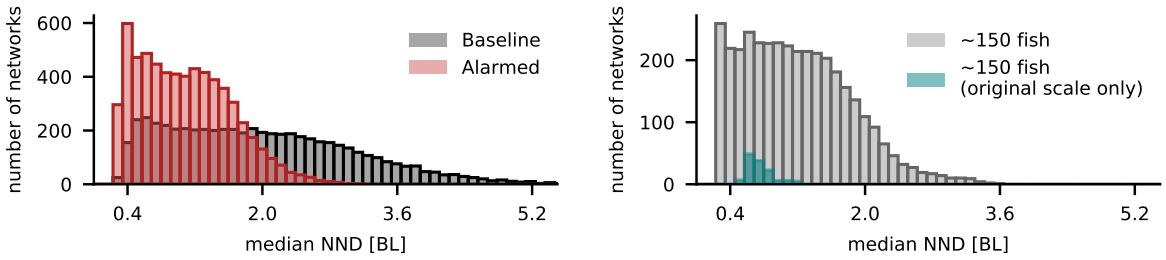

Figure S1: Histograms of median nearest neighbor distance in the rescaled position data for the different datasets. Rescaling position data (see equation (S1)) yields a wider distribution of median nearest neighbor distance in each dataset (compare distribution without rescaling, included here for 150 fish and depicted in Figure 1 in the main text for small groups).

dataset	visual field method	max. response thresh.	CI	max. LL
40 fish 'Baseline'	ray casting	0.021	[0.018,0.033]	-175
40 fish 'Baseline'	ellipses	0.028	[0.023,0.034]	-166
40 fish 'Alarmed'	ray casting	0.032	[0.030 0.034]	-387
40 fish 'Alarmed'	ellipses	0.031	[0.028,0.035]	-360
150 fish	ray casting	0.016	[0.014,0.018]	-305
150 fish	ellipses	0.019	[0.018,0.022]	-297

Table S1: Fitting the response threshold for networks constructed from the different methods of visual field construction. Credible intervals overlap for both methods and maximum likelihood is comparable. Fitting is based on 10000 simulations per network with the initial startler set in accordance with the experimentally observed cascade initiator.

dataset	β_1 (Intercept)	β_2 (LMD coefficient)	β_3 (RAA coefficient)
40 fish joined	-0.271	-2.737	-0.097
150 fish	0.302	-3.272 (-1.421)	-0.126

Table S2: Coefficient values to equation (S11) as determined by the logistic regression of first response rates. The base of the logarithm is 10 as in [28]. The values for 150 fish were taken from [27] which was using a natural log (coefficient value in bracket) and was transferred to log base 10.

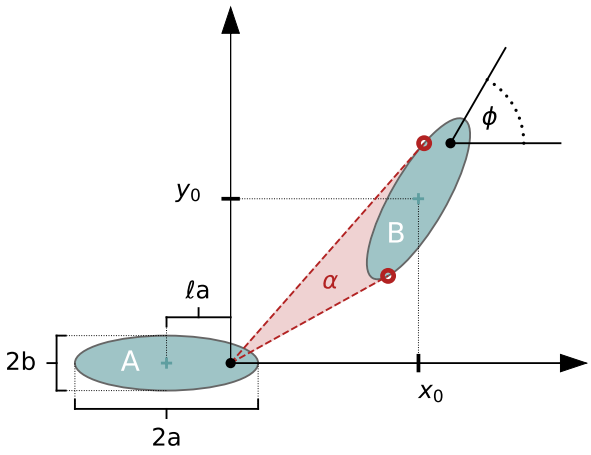

Figure S2: Illustration of the ellipses used to approximate an individual fish and variables used in the analytical calculation of the ellipse's visual field. The visual angle, α, of ellipse B in the visual field of ellipse A, is colored red and given by the angle between the two tangent lines (dashed, red) which intersect ellipse B in a single point (red) each.

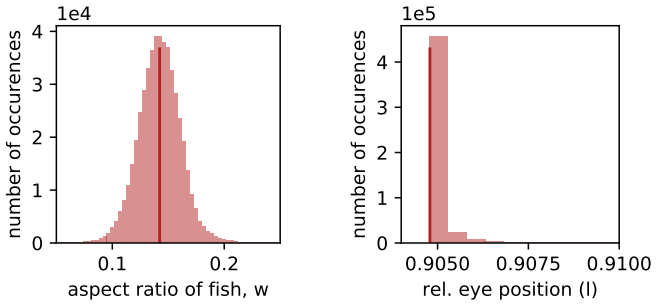

Figure S3: Determining aspect ratio w and eye position l used to approximate fish by ellipses from tracking data. This histogram is based on a video from the experiments performed in [27].

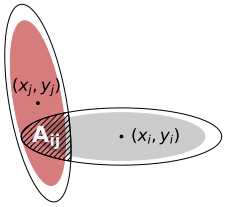

Figure S4: Illustration of the ellipse interactions in the active particle simulation. Ellipses repel each other based on their overlap area (hatched). To speed up the simulation and avoid very small forces towards the end we use a larger ellipse (rescaled by a factor 1.1, black lines) to calculate overlap area and stopped the simulation when the original sized ellipses (colored areas) do not overlap anymore.

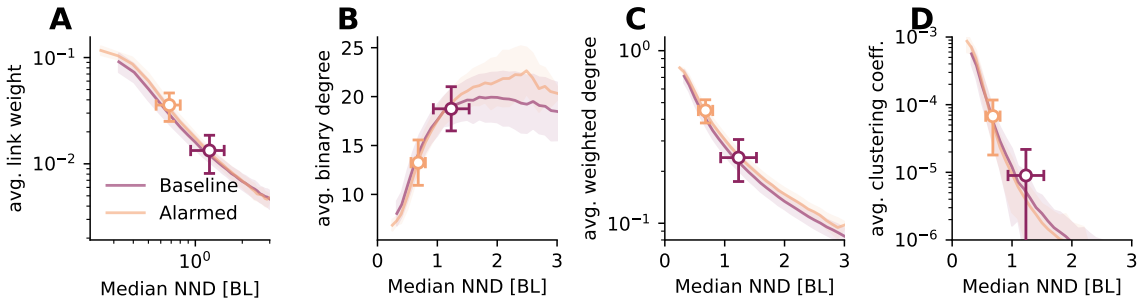

Figure S5: Network properties as a function of median nearest neighbor distance (NND) for the different datasets (lines represent averages over networks obtained from different experimental trials, shaded areas the standard deviation between networks). Dots with error bars represent averages ± one standard deviation of the mean over the original scale experimental trials. Average link weight (A) and average weighted degree (C) decrease with NND. Average number of network neighbors (B) peaks at intermediate densities. At low NND occlusions limit the number of neighbors, at high NND the visual threshold leads to a decrease in neighbors. The clustering coefficient (D), which was shown to predict cascade sizes [27], decreases with NND.

dataset	vis thresh.	max. response thresh.	CI	max. LL
40 fish 'Baseline'	0.00	0.027	[0.022, 0.033]	-164
	0.02	0.028	[0.023, 0.034]	-166
	0.10	0.049	[0.041, 0.055]	-159
	0.20	0.080	[0.056, 0.100]	-173
40 fish 'Alarmed'	0.00	0.031	[0.029, 0.035]	-360
	0.02	0.031	[0.028, 0.035]	-360
	0.10	0.039	[0.036, 0.044]	-355
	0.20	0.048	[0.041, 0.054]	-367

Table S3: Influence of a visual threshold (minimal angle required for visibility) on fitting the maximum response threshold based on 10000 simulation runs for each network. These values of the maximum response threshold are used to calculate the branching ratio, see fig. S10 and equation (S12).

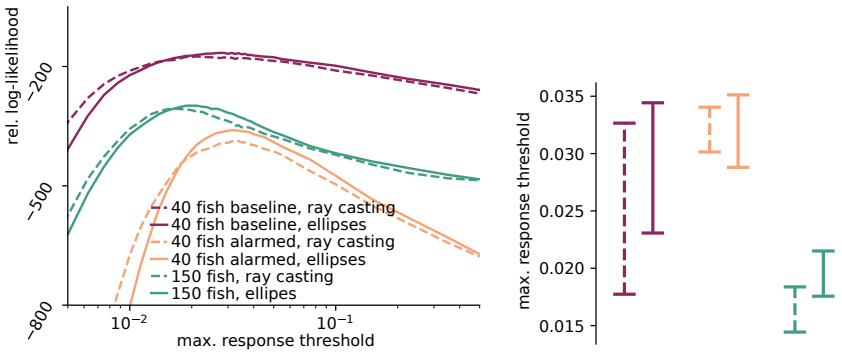

Figure S6: Calibration of the model: the maximum response threshold is chosen to maximize the rel. log-likelihood. Left: rel. log-likelihood, right: credible intervals for the best fit parameter. For details refer to table S1. While in agreement with [28] the thresholds for the two experimental conditions of the groups of 40 fish have overlapping credible intervals, the group of 150 fish is best described by a lower threshold. This difference can be due to any of the differences in experimental setup and procedure and is in need of further research exploring this systematically. The ellipse based (dashed line) and the ray casting based (solid line) networks perform comparably well in describing the experimental data.

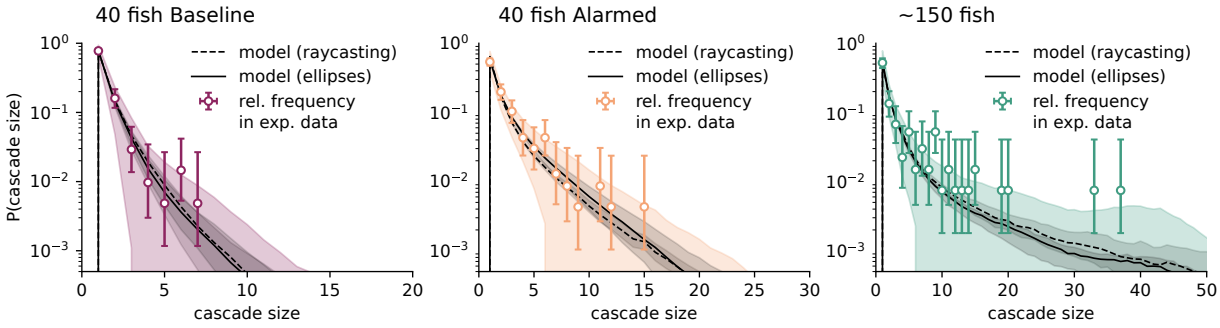

Figure S7: Best model fits (black lines) of the observed cascade size distributions. The models are using networks based on ray casting (dashed, black lines) or the ellipse approximation (solid lines) and lines represent the average over all networks. The colored (purple/orange/green) shaded areas represent +/- one standard deviation over the set of networks and therefore capture the variation in network topology and median NND. Grey shaded areas correspond to credible intervals for the model fit based on the CI of the response threshold, see Fig. S6). Dots represent relative frequencies of the different cascade sizes observed in the data as estimates of the cascade size probabilities. The error bars capture the sampling error and represent all probabilities for which the observed number of cascades lies within the 95% of a binomial distribution with this success probability. Note that the deviation of the largest two cascades in the case of large schools (right) may be due to high group polarization (see Figure S13).

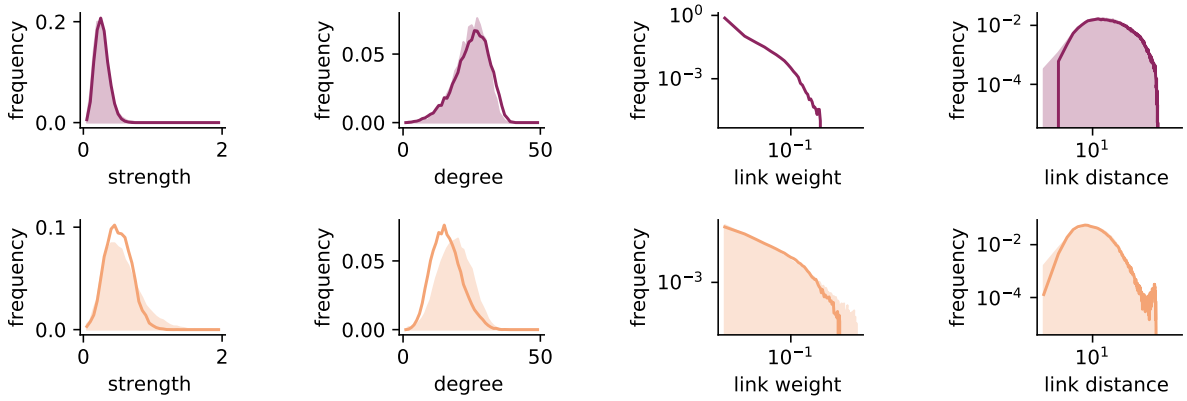

Figure S8: Comparison of networks constructed from original scale experimental data using the ellipse approximation with networks constructed from body pose estimation and ray casting (as in [28] and [27]) via distributions of different network properties. Filled area: Networks using ray casting. Line: Networks using ellipses. Histograms show (from left to right): strength (sum of all weighted network links of a node), degree (number of links per node), link weights, link distance (metric distance between two fish connected by a link). Each row corresponds to one of the datasets, top: 40 fish 'Baseline', bottom: 40 fish 'Alarmed'. Overall the approximation works well with only minor differences between the distributions. The elimination of overlaps in the ellipse networks also eliminates very small distances present in the school as seen in the right column of plots. This weakens some short and strong links by making them a bit longer (see link weight distribution). Additionally, the number of visible neighbors is decreased by this elimination of overlaps because fish that before were stacked on top of each other and not occluding each other's view now become close neighbors in the same plane and thus block a large part of each other's field of view. Especially in the case of the high density dataset (40 fish 'Alarmed') this difference is notable in the shift of the degree and strength distribution.

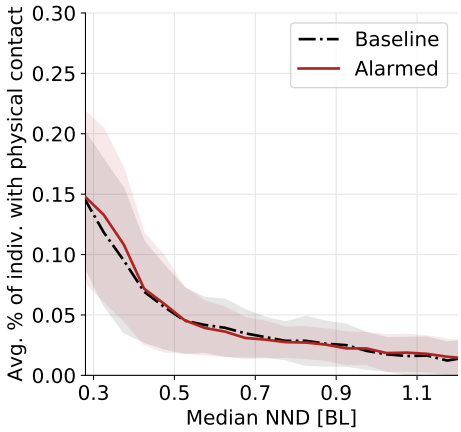

Figure S9: Investigating physical limit of density modulation: Estimation of average number of physical contacts an individual has with its neighbors at a certain density (given by the group's median nearest neighbor distance). Shaded areas indicate one standard deviation above and below the average. To estimate the number of contacts we identify ellipses that intersect when scaled up to 105% of the original size.

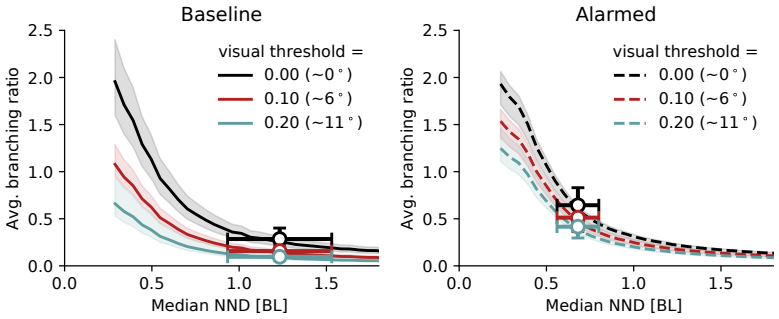

Figure S10: Average branching ratio as function of median nearest neighbor distance (NND) for different visual thresholds (minimal angle of individual i in the visual field of individual j required for existence of network link between them). We observe a shift of the estimated critical point ($b = 1$) to lower NND for increasing visual threshold, making it impossible to reach criticality without an additional change of individual responsiveness. This does not change our finding, that experimentally observed schools are subcritical, nor the general form of the relative individual payoff.

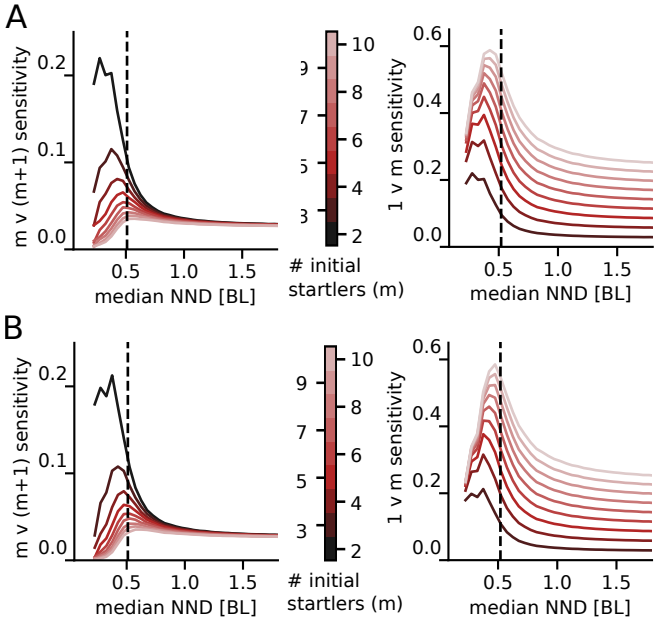

Figure S11: Alternative definitions of collective sensitivity as difference in average cascade size between m and $m+1$ (left) or 1 and m (right) initial startlers all exhibit a peak near criticality. A) Initial startlers are chosen randomly from the school. B) Initial startlers are randomly chosen network neighbors of the same (randomly chosen) individual. Results are qualitatively similar for both choices of initialization. The dashed vertical line indicates where $b = 1$. For all measures, the location of peak sensitivity is near this line and its distance to it decreases with increasing m.

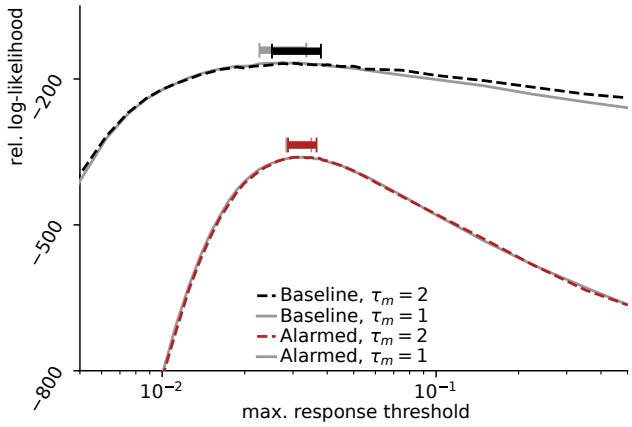

Figure S12: Relative log-likelihood for two different values of agent memory: $\tau_m = 1$ (dashed lines, as used in main text) and $\tau_m = 2$ (solid lines, as used in [28]. The optimal value of the max. response threshold does not change significantly for the two choices.

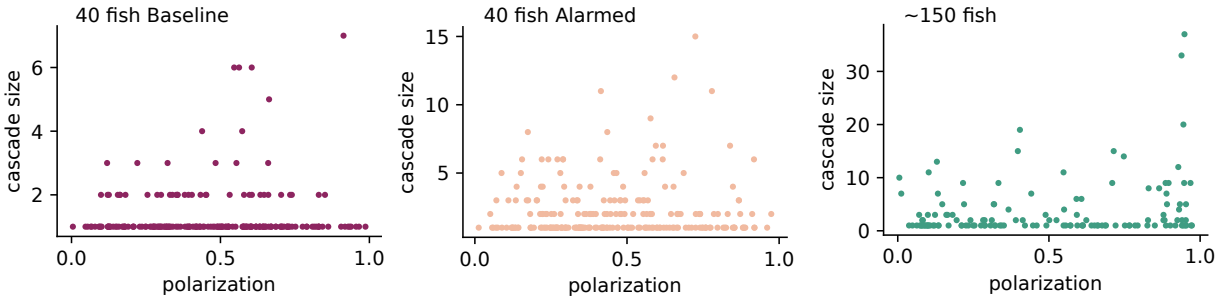

Figure S13: Observed cascade sizes plotted against group polarization (see equation (S13)) at the beginning of the cascade for the three different datasets.

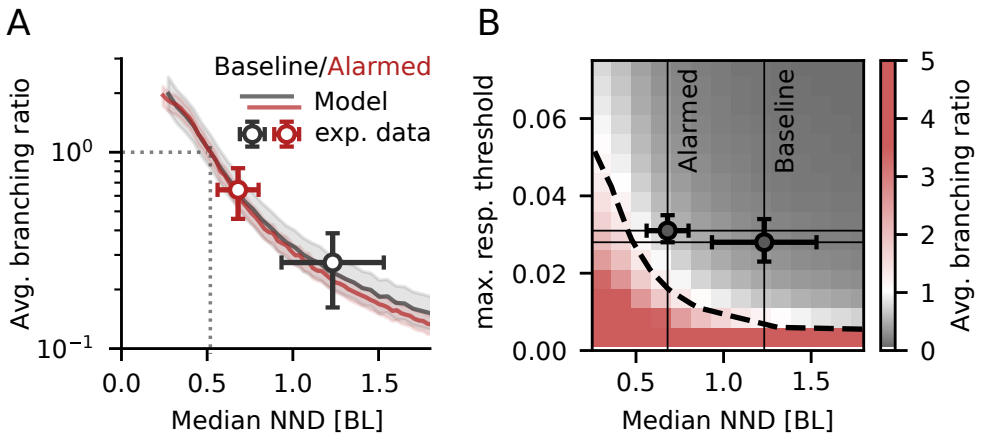

Figure S14: The average branching ratio, an analytical estimate of criticality, as a function of A) median nearest neighbor distance (NND) and B) median NND and maximum response threshold. Lines in A) are averages over simulations for all rescaled networks binned by median NND. Shaded areas indicate the uncertainty of the model fit (average response threshold fit, see Table S1). The dotted line in A) and dashed line in B) marks $b = 1$ which is also include in Fig. 3C of the main text. For the optimal response threshold both datasets indicate a critical median NND of 0.52 BL. Data points are averages over original scale networks and represent the experimentally observed schools. In A) both horizontal and vertical error bars indicate one standard deviation of the average over networks in B) the vertical errorbars are the credible intervals for the model fit of the average response threshold, see table S1. For version of A) including the scattered data points of original scale networks instead of their average, refer to Figure S25.

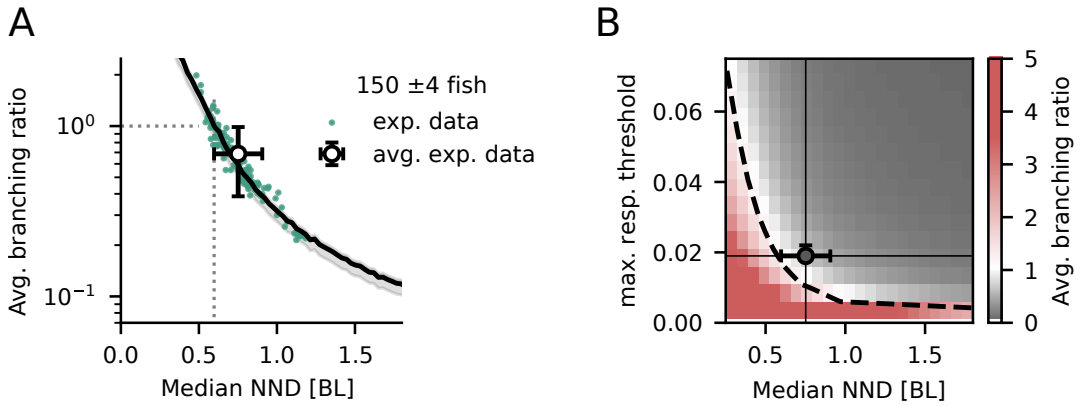

Figure S15: Branching ratio for dataset of approximately 150 fish from [27]. Analytical estimation averaged over small bins of median NND using the full set including rescaled networks (lines in A, colormap in B) and averages over only original scale networks characterizing experimental observations (errobars, as in S14). For the optimal response threshold ($\theta_{\max} = 0.19$) the critical median NND is approximately 0.6 BL (dotted vertical line in A). The observed schools are on average subcritical, like the observed schools of 40 fish discussed in the main paper, but some single observed schools are have $b > 1$. Shaded grey areas indicate the branching ratio resulting from using the upper and lower limit of the credible interval of the average response threshold. The average nearest neighbor distance of 0.75±0.15 BL) is comparable to that of the 'Alarmed' dataset of 40 fish, but the lower fitted average response threshold for this dataset yields larger branching ratios in comparison.

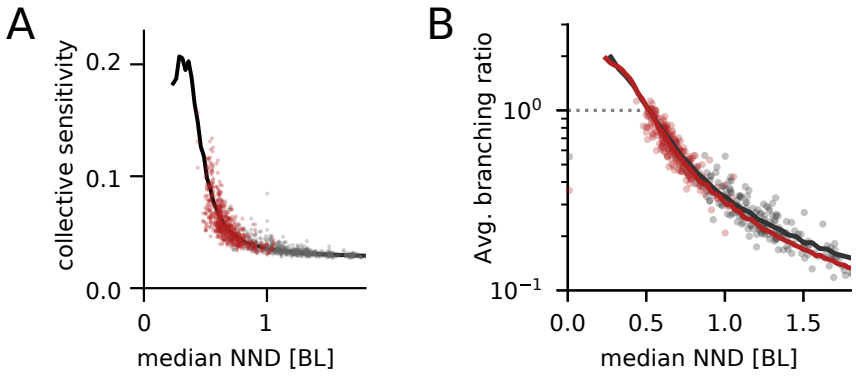

Figure S16: Variation of criticality measures within the model prediction for the experimentally observed schools. Each point represents one of the original scale networks from either the 'Alarmed' (red) or the 'Baseline' dataset (grey). The 'Alarmed' dataset contains a few networks with a critical branching ratio but collective sensitivity below the maximal value.

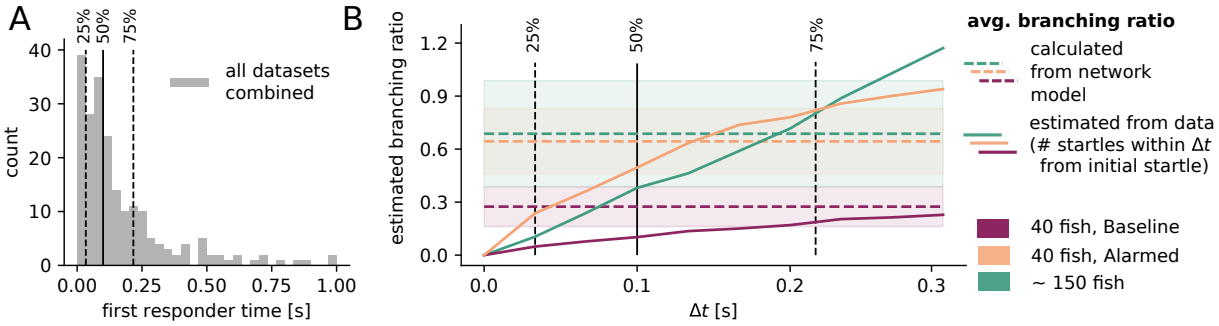

Figure S17: Estimation of average branching ratio directly from data of observed cascades. (A) Distribution of first responder times (time between first and second observed startle) across all three datasets. Vertical lines indicate quartiles of the distribution. (B) Branching ratio estimates based on the average number of startles within a time interval Δt after the initial startle (solid lines) depend strongly on Δt. For $0.1\,\mathrm{s} \leq \Delta t \leq 0.22\,\mathrm{s}$ (a choice of Δt which includes between 50% and 75% of all observed initial startles) the estimates of the branching ratio from the measured activity (solid lines) lie within one standard deviation of the branching ratio estimate based on the reconstructed networks used in the main text (dashed lines). Shaded areas indicate one standard deviation of the average over networks for the network-based branching ratio estimate.

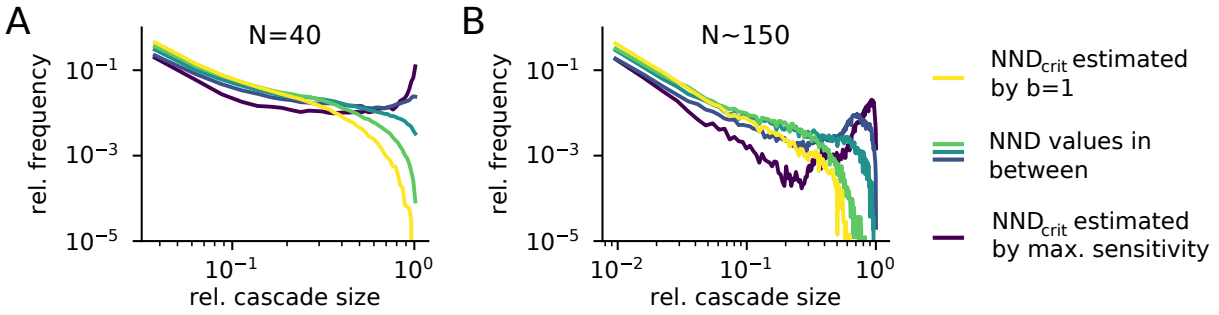

Figure S18: Cascade size distributions for $\theta_{\max} = 0.016$. The two estimates of criticality, the critical branching ratio, $b = 1$, and the maximum sensitivity, provide an upper and lower bound for the median nearest neighbor distance at which the cascade size distribution most resembles a truncated power law, both for the smaller (N=40, A) and the larger system ($N \approx 150$, B). Lines represent distributions of simulation runs for networks (1000 runs per network) within an (approx.) 0.05 BL interval centered around estimated critical median NNDs (yellow: $b = 1$, $\mathrm{NND}_{\mathrm{crit}} = 0.65\,\mathrm{BL}$ for $N = 40$ and $\mathrm{NND}_{\mathrm{crit}} = 0.74\,\mathrm{BL}$ for $N \approx 150$, dark blue: max. sensitivity, $\mathrm{NND}_{\mathrm{crit}} = 0.50\,\mathrm{BL}$ for $N = 40$ and $\mathrm{NND}_{\mathrm{crit}} = 0.45\,\mathrm{BL}$ for $N \approx 150$) and equidistant NND values inbetween the two estimates.

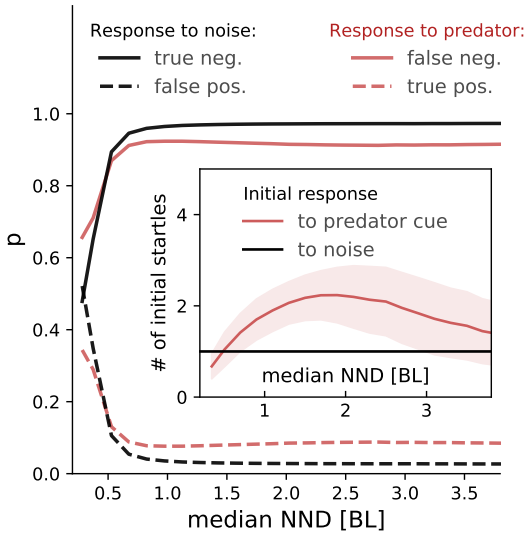

Figure S19: Likelihood of the different decision outcomes for the individual as a function of median nearest neighbor distance (NND). Close to the critical point false negatives decrease and false positives increase, and thus there remains a trade-off between two types of errors that can be managed according to the environment by choosing the appropriate distance to criticality. Inset shows the number of initial startlers used to trigger cascades assumed to be initiated by a predator cue (red) or noise (black).

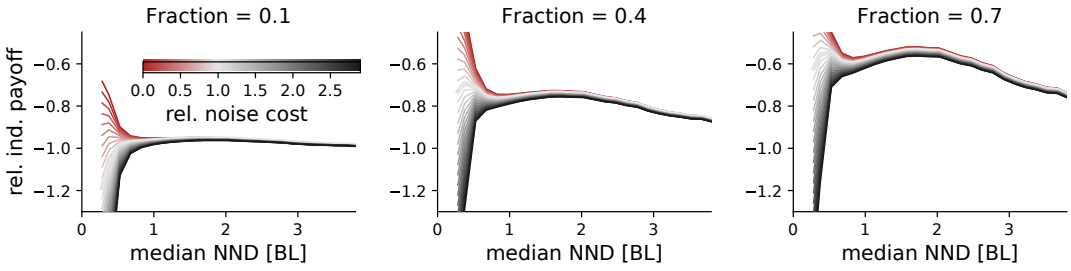

Figure S20: Relative payoff for different reactive fractions, p_{detect}, in the individuals that can see the predator. Increasing this fraction increases the maximum at intermediate densities but does not change the qualitative results. The left panel is as in the main paper.

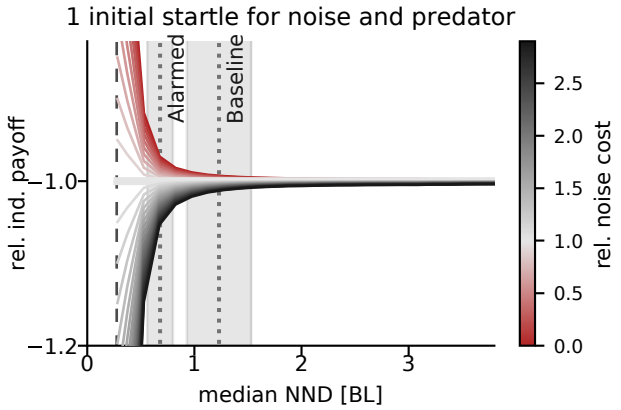

Figure S21: Relative payoff, assuming that both noise and predator cue trigger just one initial startle. The divergence at small NND remains, while the maximum at intermediate densities and the one at the critical point disappear.

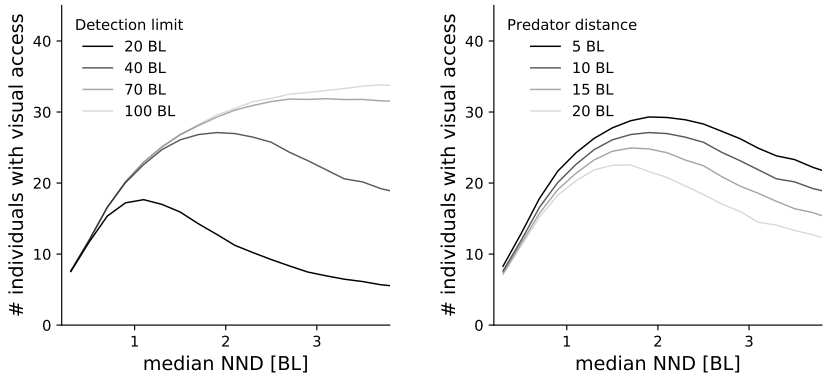

Figure S22: Influence of distance of group boundary to predator (d_{pred}, right) and maximal detection distance of the individual (d_{max}, left) on the number of individuals that have visual access to a predator plotted against median nearest neighbor distance. As long as there is an upper limit to the distance at which an individual can perceive a predator, the qualitative shape of the curve remains unchanged. The exact position of the maximum changes.

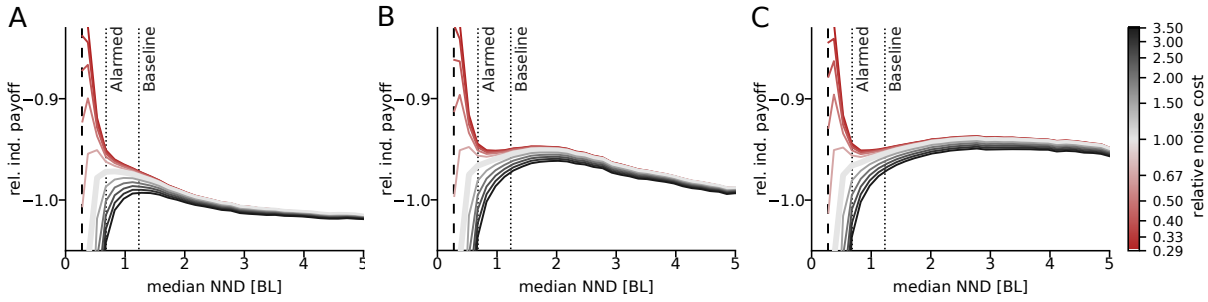

Figure S23: Relative individual payoff for varying parameter $d_{\max}$ (maximum detection limit) of the visual detection for $d_{\max} = 20\,\mathrm{BL}$ (left), $d_{\max} = 40\,\mathrm{BL}$ (middle, as in main text) and $d_{\max} = 70\,\mathrm{BL}$ (right). All plots use $d_{\mathrm{pred}} = 10\,\mathrm{BL}$. Due to the shift in the position of the maximum of visual access (see Fig. S22, left plot) the position of the second maximum of the relative payoff also shifts. For $d_{\max} = 20\,\mathrm{BL}$ (left) both maxima (the criticality-based one and the visual-access-based one) merge into a single maximum. For this case, the average median NND of the 'Baseline' dataset is optimal for high relative noise costs ($\xi \approx 3$) and the 'Alarmed' dataset is on average optimal for a relative noise cost of $\xi \approx 1$.

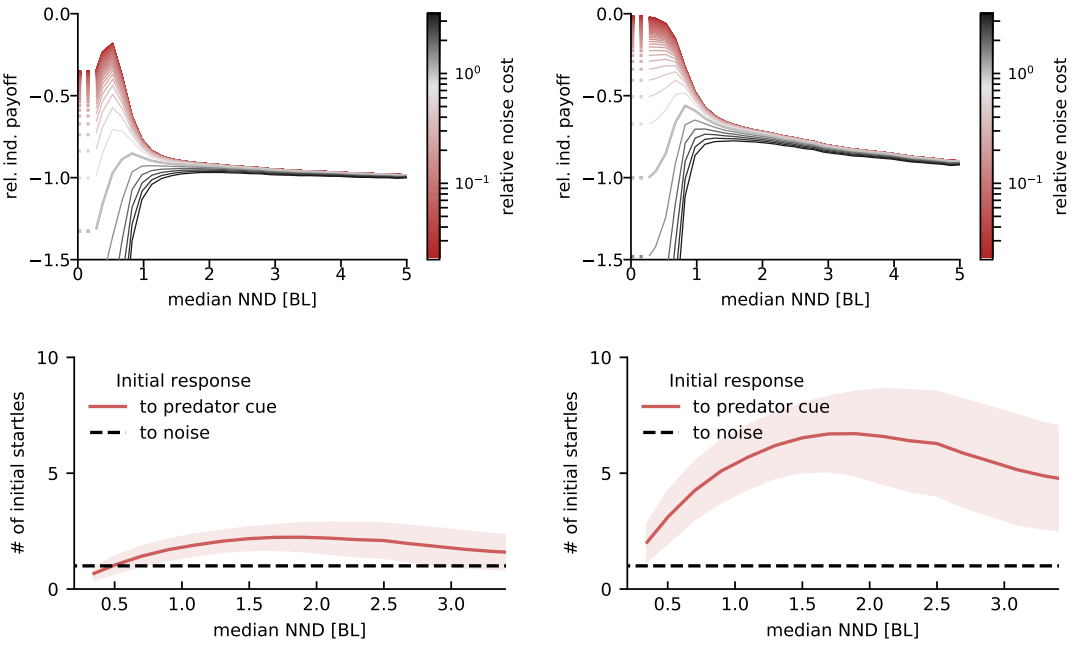

Figure S24: Payoff for an maximum response threshold of $\theta_{\max} = 0.009$ and reactive fraction $p_{\text{detect}} = 0.1$ (left) or $p_{\text{detect}} = 0.3$ (right). Because at low NND the number of initial startlers goes below 1 for $p_{\text{detect}} = 0.1$, the supercritical state (very low NND), is not optimal. Even though every initial startle yields a global response in this regime, the likelihood of a predator being detected is lower than one and thus no cascade may be initiated at all. This decrease in payoff for very low NND and very low relative noise cost (red curves) disappears as soon as we assume that at least one individual will respond to a predator (see right plots). The existence of a maximum at criticality for intermediate relative noise cost is unchanged by this. Unlike for $\theta_{\max} \approx 0.03$ as used in the main text, here the second maximum disappears because the critical point has shifted to higher NND and thus closer to the maximum of the visual detection (bottom plots).

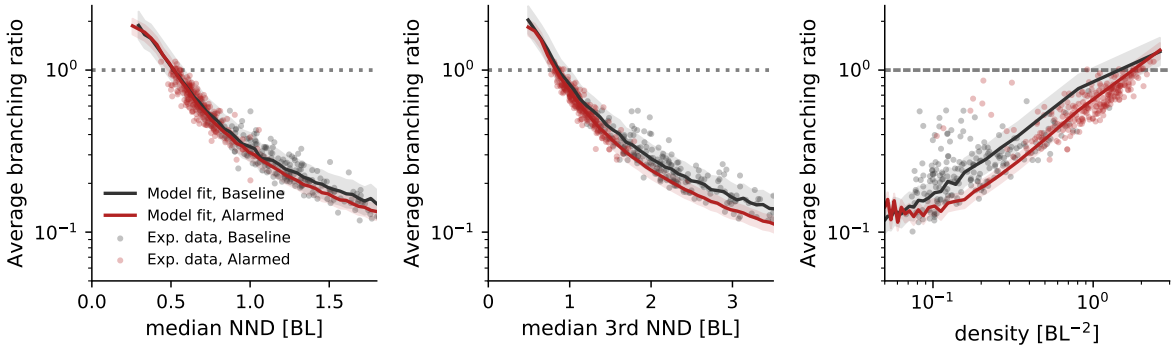

Figure S25: Average branching ratio plotted against three measures of density: median nearest neighbor distance (left, as used in main text), third nearest neighbor distance (middle) and density (right, calculated as ratio of group size and area of the group's convex hull). Data points represent results using the networks obtained from experimental observations without rescaling of inter-individual distances. Independent of the density measure the observed schools are on average subcritical with single networks having a critical branching ratio, $b = 1$.