

*The content and availability of information
affects the evolution of social-information
gathering strategies*

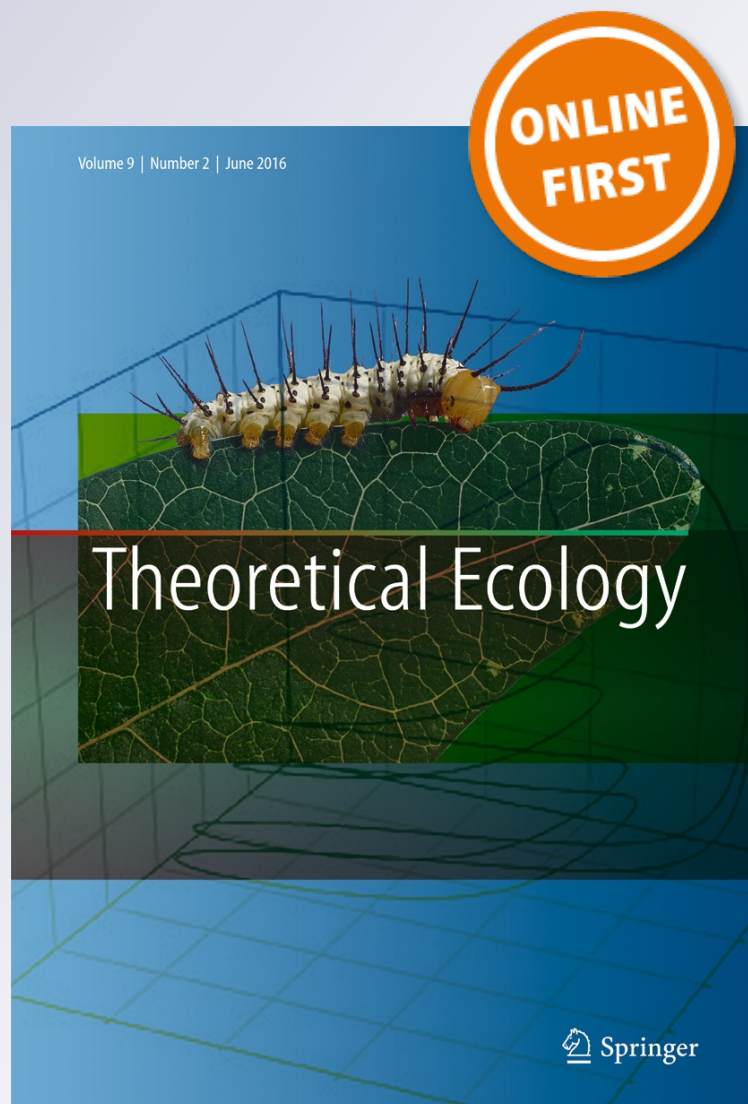
**Eleanor Redstart Brush, Naomi Ehrich
Leonard & Simon A. Levin**

Theoretical Ecology

ISSN 1874-1738

Theor Ecol

DOI 10.1007/s12080-016-0301-4



Your article is protected by copyright and all rights are held exclusively by Springer Science +Business Media Dordrecht. This e-offprint is for personal use only and shall not be self-archived in electronic repositories. If you wish to self-archive your article, please use the accepted manuscript version for posting on your own website. You may further deposit the accepted manuscript version in any repository, provided it is only made publicly available 12 months after official publication or later and provided acknowledgement is given to the original source of publication and a link is inserted to the published article on Springer's website. The link must be accompanied by the following text: "The final publication is available at link.springer.com".

The content and availability of information affects the evolution of social-information gathering strategies

Eleanor Redstart Brush¹ · Naomi Ehrich Leonard² · Simon A. Levin³

Received: 3 January 2016 / Accepted: 8 May 2016
© Springer Science+Business Media Dordrecht 2016

Abstract Social animals can gather information by observing the other members of their groups. Strategies for gathering this type of social information have many components. In particular, an animal can vary the number of other animals it observes. European starlings (*Sturnus vulgaris*) in flight pay attention to a number of neighbors that allows the flock to reach consensus quickly and robustly. The birds may do this because being in such a flock confers benefits on its members, or the birds may use the strategy that is individually beneficial without regard for the flock's structure. To understand when individual-level optimization results in a group-level optimum, we develop a model of animals gathering social information about environmental cues, where the cue can be about either predators or resources, and we analyze two processes through which the number of neighbors changes over time. We then identify the number of neighbors the birds use when the two dynamics reach equilibrium. First, we find that the equilibrium number of neighbors is much lower when the birds are learning about the presence of resources rather than predators. Second, when the information is about the presence of predators, we find that the equilibrium number of neighbors increases as the

information becomes more widespread. Third, we find that an optimization process converges on strategies that allow the flock to reach consensus when the information is about the presence of abundant resources, but not when it is about the presence of scarce resources or predators.

Keywords Evolution · Consensus · Correlation length · Information · Robustness · Social

Introduction

In animal social groups, what is best for the group is not always what is best for any given animal. One of the challenges of evolutionary theory is to understand when and why an animal contributes to the good of its social group when that behavior is costly. The persistence of cooperation, for instance, is puzzling when an individual has an incentive to avoid paying the costs associated with cooperative behavior. The use of social information in animal groups is another such puzzle. If a group can reach consensus about where to go and what to do at any particular moment, its members can experience safety in numbers (Landeau and Terborgh 1986; Molvar and Bowyer 1994) and greater foraging efficiency (Sullivan 1984; Berger 1978; Pitcher et al. 1982), among other benefits. However, reaching consensus may require that each animal sacrifice its own ability to explore or to move in its preferred direction. In this paper, we build and analyze a mathematical model of a group of animals using social information to study how the conflict between the group's ability to reach consensus and the individuals' efforts to learn about the environment can be reconciled.

Animals observe their surroundings and use that direct personal information to decide where to go next. Animals

✉ Eleanor Redstart Brush
eleanor.brush@gmail.com

¹ Program in Quantitative and Computational Biology,
Princeton University, Princeton NJ, USA

² Department of Mechanical and Aerospace Engineering,
Princeton University, Princeton NJ, USA

³ Department of Ecology and Evolutionary Biology, Princeton
University, Princeton NJ, USA

in groups can also use the inadvertent social information (sensu, Dall et al. 2005; Danchin et al. 2004) produced by the movement of their peers when making these decisions. Animals have been shown to be influenced by the movements of their group-mates in species of many taxonomic groups, including fish (Katz et al. 2011; Strandburg-Peshkin et al. 2013), birds (Bialek et al. 2012; Bialek et al. 2014; Cavagna et al. 2009; Porter and Sealy 1982; Weimerskirch et al. 2010), bats (Dechmann et al. 2009), and primates (Strandburg-Peshkin et al. 2015). Groups of animals responding to each other's behavior can produce spectacular collective dynamics, like the well-known examples of starling murmurations and sardine schools. Information about both the location of food (Porter and Sealy 1982; Strandburg-Peshkin et al. 2013; Weimerskirch et al. 2010) and the presence of predators (Attanasi et al. 2014; Powell 1974; Rosenthal et al. 2015) can spread through such groups, even when only a few members have direct personal information. If the members of a population are all gathering inadvertent social information from their peers, an animal's strategy for choosing whom to learn from can also affect the animals who learn from it. This makes social-information use interesting from an evolutionary perspective because, rather than there being a universally optimal strategy, the optimal strategy depends on the social context set by what the rest of the group is doing.

In large schools of fish and flocks of birds, there are probably simple rules, based on spatial proximity, that determine from whom an animal gathers information. An animal might only pay attention to others within some distance of itself (within a "metric" neighborhood around itself), it might pay attention to a fixed number of animals (within a "topological" neighborhood), or it might use a combination of these strategies, for example by paying attention to a fixed number of animals provided they are within some distance. Scientists have used the alignment or correlation between the velocities of animals in moving groups to study which of these rules, if either, determine to whom an animal pays attention. A metric rule successfully describes the schooling dynamics of golden shiners (*Notemigonus crysoleucas*) (Katz et al. 2011), although taking the visual field into account improves predictions of how information spreads across schools of these fish (Strandburg-Peshkin et al. 2013). Surf scoters (*Melanitta perspicillata*) also seem to flock using metric rules (Lukeman et al. 2010). Flocks of European starlings (*Sturnus vulgaris*) have been described using both metric (Bode et al. 2010) and topological rules (Ballerini et al. 2008; Bialek et al. 2012). Among scientists using topological rules to describe starling flocks, Ballerini et al. (2008) found that starlings pay attention to their six or seven nearest neighbors, while Bialek et al. (2012) argued that ten is the maximum-likelihood estimate of the number of neighbors the birds use.

If a topological rule determines how starlings gather information, the number of neighbors to which a bird pays attention affects the information it gathers. It also affects the flock's ability to reach consensus. Two measures of this ability have been applied to data collected from starling flocks in flight. One, called \mathcal{H}_2 robustness, describes how close to consensus the birds in a flock can stay, despite errors and noise. Young et al. (2013) used the positions of starlings in several flocks to construct interaction networks using various numbers of nearest neighbors. They found that the networks constructed using seven nearest neighbors had the highest robustness, which matches the number Ballerini et al. (2008) argued the birds are actually using. The other measure of a flock's ability to reach consensus is its correlation length, a concept borrowed from statistical mechanics. Roughly, the correlation length of a flock is the distance at which birds' opinions are uncorrelated. The correlation length of starling flocks scales linearly with the number of birds in the flock, which suggests that the flocks are close to criticality (Cavagna et al. 2009; Bialek et al. 2014). Criticality can enable a system to respond optimally to changes in the environment: not so slowly as to never respond, nor so quickly as to be unstable (Bak et al. 1988). This is another piece of evidence that starlings construct interaction networks with optimal structure.

There are two possible explanations for the observations that flying starlings construct interaction networks that both maximize robustness and exhibit criticality. The first is that these group-level properties provide direct benefits to the birds, and that, through either evolution or some learning process, the birds acquire strategies that optimize these properties. The second is that a bird uses the strategy that optimizes its own ability to gather information from its peers, and that optimal group-level properties emerge as a consequence of this individual-level optimization. Here, we analyze an evolutionary model of social-information use in starling flocks, in order to study (1) how the evolution of the number of neighbors the birds pay attention to depends on whether the birds are trying to gather information about predators or resources and (2) whether the strategies that evolve promote or diminish the performance of the flock as a whole. While starlings surely gather inadvertent social information while resting and foraging, we focus on their time spent flying, as this is the behavior that was analyzed in the studies we summarized above. Our model was inspired by starling flocks, but it is general enough that our results should apply to other species in which topological rules determine their collective dynamics. While there are many models of the evolution of social-information use, few, to our knowledge, compare information about different facets of the environment (in our case, predators and resources). Additionally, while some models of the evolution of social-information use consider the weight given

to social information (as opposed to personal information) (e.g., Pais and Leonard 2014; Shaw and Couzin 2013; Torney et al. 2015), none to our knowledge consider the evolution of the number of neighbors in a topological interaction network. Finally, our work considers two group-level properties that have not previously been studied together.

Model

How information about a cue spreads

Each bird has an opinion about the environment, which it changes based on observations of other members of the flock. Periodically, an environmental cue appears: there is either a predator that should be avoided or resources that should be consumed. This can happen because a predator approaches the flock, because new resources appear, or because the flock moves to a new area in which either predators or resources are present. We assume that the cues appear relatively rarely compared to the timescale on which the birds in a flock move and rearrange themselves. The environment and the type of cue will affect how many members of the flock perceive the cue. For example, if the flock is flying through a dense city and a small hawk approaches, only a few starlings may notice. On the other hand, if the flock is flying over an open field covered in corn, all of the birds will perceive that there is food to be had. The birds that perceive the cue change their opinions based on the information they receive, while continuing to pay attention to other birds in the flock.

Specifically, when a cue appears, we assign all birds a new position by drawing positions randomly from the unit square. Each bird has an opinion about the environment that can take any real value, $v_i \in [0, 1]$. (Table 1 lists the variables used in the text.) A bird's opinion can be interpreted as its direction of flight and having an opinion equal to 1 can be interpreted as responding correctly to the cue, i.e., flying directly away from a predator or toward resources. Each bird i observes the k_i birds nearest to it, which form i 's neighborhood \mathcal{N}_i . To determine which birds perceive the cue, we pick a focal bird at random. Then, all birds within a radius r of the focal bird perceive the cue. If a bird does not perceive the cue, it changes its opinion to be close to the average opinion of its neighbors:

$$\frac{dv_i}{dt} = \frac{1}{k_i} \sum_{j \in \mathcal{N}_i} (v_j - v_i).$$

If a bird perceives the cue, it also changes its opinion to be close to 1:

$$\frac{dv_i}{dt} = \frac{1}{k_i} \sum_{j \in \mathcal{N}_i} (v_j - v_i) + b(1 - v_i),$$

Table 1 Table of matrices and variables used in the text

A	weighted adjacency matrix	$A_{ij} = \begin{cases} \frac{1}{k_i} & \text{if } j \in \mathcal{N}_i \\ 0 & \text{otherwise} \end{cases}$
B	diagonal cue matrix	$B_{ii} = \begin{cases} b & \text{if } i \text{ perceives cue} \\ 0 & \text{otherwise} \end{cases}$
I_n	$n \times n$ identity matrix	
λ	correlation length	
L	negative of Laplacian w/o cue	$L = A - I$
L_s	negative of Laplacian w/ cue	$L_s = L - B$
M	interaction matrix	
\bar{M}	M rotated away from consensus	$\bar{M} = QMQ^T$
\tilde{M}	M rotated into consensus basis	$\tilde{M} = RMR^T$
P	symmetrized version of L	$P_{ij} = \begin{cases} S_{ij} & \text{for } j \neq i \\ -\sum_k S_{ik} & \text{for } j = i \end{cases}$
P_s	symmetrized version of L_s	$P_s = P - B$
Q	$Q^T Q$ projects away from $\mathbf{1}$	$Q \in \mathbb{R}^{(N-1) \times N}, Q\mathbf{1} = \mathbf{0}$
R	rotation matrix	$R \in \mathbb{R}^{N \times N}, R = \begin{pmatrix} Q \\ \frac{1}{\sqrt{N}}\mathbf{1}^T \end{pmatrix}$
ρ	\mathcal{H}_2 robustness	$y = Qv, \rho = \lim_{t \rightarrow \infty} \frac{1}{\sqrt{E[\ y\ ^2]}}$
$\bar{\rho}$	average ρ	
S	symmetrized version of A	$S = \frac{1}{2}(A + A^T)$
v	opinion vector	

where b is a parameter that describes the importance of the environmental cue. An example of these opinion dynamics is shown in Fig. 1.

The flock will eventually reach consensus with every $v_i = 1$, as long as there is a path in the network from every node to at least one node that perceives the cue (Pais and Leonard 2014). In the Appendix A.2, we show that, if bird i does not perceive the cue, then

$$v_i(t) \approx v_i(0) + \frac{bp_i t^2}{2},$$

where p_i is the proportion of i 's neighbors that perceive the cue. We assume that all birds have $v_i(0) = \frac{1}{2}$ when the cue appears, so, if bird i does not perceive the cue, then

$$v_i(t) \approx \frac{1}{2} \left(1 + t^2 bp_i \right). \tag{1}$$

It will be useful to write the system of differential equations more compactly. If

$$A_{ij} = \begin{cases} \frac{1}{k_i} & \text{if } j \text{ is in } \mathcal{N}_i \\ 0 & \text{otherwise} \end{cases} \quad \text{and} \quad L_{ij} = \begin{cases} A_{ij} & \text{if } j \neq i \\ -1 & \text{if } j = i \end{cases},$$

then L is the negative of the random-walk normalized Laplacian of the network of interactions. If B is the diagonal matrix with $B_{ii} = b$ if i perceives the cue and $B_{ii} = 0$ otherwise, and $L_s = L - B$, then L_s can be used to write

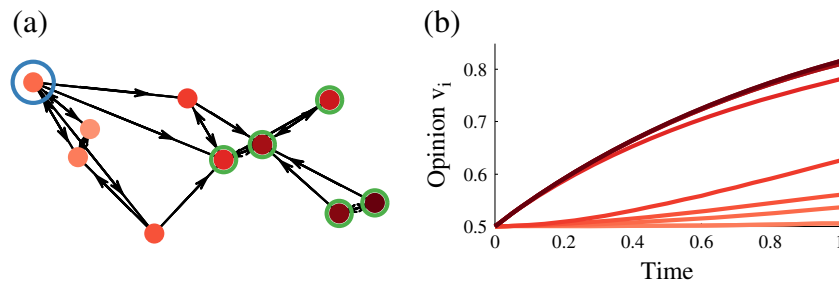


Fig. 1 An example of the opinion dynamics after a cue. In **a**, we show a group of $N = 10$ birds, all of which have two neighbors except for the *bird circled in blue*, which has five neighbors. The *birds circled in green* perceive the cue. An *arrow points* from i to j if i pays attention to j . In **b**, we show the opinion dynamics, given this adjacency matrix

and the birds that perceive the cue. Birds with similar positions in the network have similar opinion dynamics, so that some of the birds' lines overlap each other. The colors of the nodes in **(a)** and lines in **(b)** indicate the opinion of each bird at $t = 1$, $v_i(1)$. Parameters: $N = 10$, $r = 0.4$, $b = 1$

the system of differential equations in terms of the vector of opinions $v = (v_1, \dots, v_N)^T$:

$$\frac{dv}{dt} = Lv + B(\mathbf{1} - v) = L_s v + B\mathbf{1},$$

where $\mathbf{1} = (1, \dots, 1)^T$.

Fitness

Each bird's fitness is determined by how well-informed it is, i.e., how close its opinion is to 1. If the cue is about a predator, the least well-informed bird will be eaten. If the cue is about resources, the most well-informed birds will get to eat. In each case, we assume the critical event—birds being eaten or getting to eat—occurs one unit of time after the cue appears.

Since how well-informed a bird is depends on the interaction network and on which birds perceive the cue, we simulate 1000 cues to determine the fitness of each individual in a flock whose strategies are $\{k_1, \dots, k_N\}$. If the cue provides information that a predator is present,

$$P(i \text{ survives}) = 1 - P(i \text{ is eaten}) \\ = 1 - \frac{\# \text{ of cues such that } i = \operatorname{argmin}_j \{v_j(1)\}}{1000}$$

and i 's fitness is proportional to this probability. If the cue provides information about scarce resources,

$$P(i \text{ gets to eat}) = \frac{\# \text{ of cues such that } i = \operatorname{argmax}_j \{v_j(1)\}}{1000}$$

and i 's fitness is proportional to this probability. If the cue provides information about abundant resources, the probability that bird i gets to eat is the probability that $v_i(1)$ is one of the four highest opinions over all cues and i 's fitness is proportional to this probability. (Note that abundant is a relative term. We consider a flock of size $N = 20$ so that only a fifth of the flock gets access to resources. Sensitivity to the number of birds that are allowed to eat is discussed in Fig. 11.) Finally, we consider a combination of predation

and access to scarce resources. To do well, a bird must both survive and get to eat. In this case, i 's fitness is proportional to $P(i \text{ survives}) \times P(i \text{ gets to eat})$. In each of the probabilities defined above, if any set of birds has exactly the same opinion, then one of them is picked at random to be the one that is eaten or gets to eat.

In the opinion dynamics we defined above, we assumed a linear response function: a bird's direction of flight was exactly equal to its opinion about the environment. However, there are other possible response functions. For example, if a bird only responds to the cue when its opinion passes some threshold, its response follows a step function. Regardless of its specific functional form, any reasonable response function should be monotonic in the bird's opinion. Since any monotonic transformation of the flock's opinions will preserve the argmin and argmax of the flock, changing the response function would not affect our definitions of fitness or our results.

Strategy dynamics

The number of neighbors a bird uses may be a genetically inherited trait or a learned strategy. In either case, the frequencies of strategies that confer higher fitness should increase over time, until no bird can improve its fitness by changing its strategy. These equilibrium strategies are a reasonable prediction for what we would expect to find in nature. We use two different methods for analyzing the dynamics of the number of neighbors the birds use. First, we identify the evolutionarily stable strategy. The ESS is the strategy k that no other strategy k' can invade, where we assume that a strategy k' invades a group using k if the fitness of individuals using k' is higher than the fitness of those using k . Equivalently, if all members of the group use the ESS, no individual has an incentive to use another strategy. Details on how to find the ESS in a finite group are provided in Appendix A.1. Assuming all members of the group use the same strategy allows us to exhaustively search over

all strategies to find the ESS. However, it is a rather strict constraint. Our second method of optimization allows us to consider how the birds might optimize their strategies in a heterogeneous group. Given a random set of initial strategies, we allow each bird to choose the strategy that would give it the highest fitness, given the rest of the strategies being used. We repeat this process until the birds reach an equilibrium set of strategies. We repeat this over many initial sets of strategies to find average properties of this greedy optimization process.

Measures of how well the flock reaches consensus

We consider two ways of measuring how well the flock can reach consensus. The first, \mathcal{H}_2 robustness, captures how close the flock's opinions are to being in consensus. The second, correlation length, captures how far information can travel through the flock. When information can travel far through the flock, it is usually the case that it is easy for the flock to stay in consensus. However, it can happen that the flock has a low robustness even though it has a high correlation length and vice versa.

To define \mathcal{H}_2 robustness, we have to measure how close the flock's opinions are to being in consensus. We therefore transform the opinion vector v into a vector y whose length is the amount of "disagreement" in the flock, in the language of Young et al. (2013). Following Young et al. (2010, 2013), we define $Q \in \mathbb{R}^{(N-1) \times N}$ such that each row is normal and orthogonal to the consensus vector $\mathbf{1}$. This is equivalent to finding Q such that $Q^T Q = I_N - \frac{1}{N} \mathbf{1}\mathbf{1}^T$ and $Q Q^T = I_{N-1}$ (Young et al. 2010, 2013). If we define $y = Qv$, then y is the component of v that is orthogonal to the consensus vector and $\|y\|$ describes how close the flock's opinion are to being in consensus (Young et al. 2010, 2013). This allows us to define \mathcal{H}_2 robustness (Bamieh et al. 2012; Young et al. 2010, 2013; Zelazo and Mesbahi 2009):

$$\rho = \lim_{t \rightarrow \infty} \frac{1}{\sqrt{E[\|y(t)\|^2]}}$$

If we define the covariance matrix

$$\Sigma_{yij}(t) = E[(y_i(t) - E[y_i(t)])(y_j(t) - E[y_j(t)])]$$

and the steady-state covariance matrix $\Sigma_{ss,y} = \lim_{t \rightarrow \infty} \Sigma_y(t)$ then

$$\rho = \frac{1}{\sqrt{\text{Tr}(\Sigma_{ss,y})}}$$
 (Young et al. 2010, 2013).

If it is easy for the flock to reach consensus, y will tend to be small and ρ will be large. Given any set of strategies, we find the average of ρ over 1000 cues, $\bar{\rho}$, since ρ depends on the specific interaction network.

To define correlation length, we find, for each pair of birds, the correlations between the deviations in their opinions. We now transform v by defining $\bar{v} = \sum_i v_i/N$ and $z = v - \bar{v}\mathbf{1}$. Note that $z = Q^T Qv = Q^T y$. If we define the covariance matrix

$$\Sigma_{zij}(t) = E[(z_i(t) - E[z_i(t)])(z_j(t) - E[z_j(t)])]$$

and the steady-state covariance matrix $\Sigma_{ss,z} = \lim_{t \rightarrow \infty} \Sigma_z(t)$, then the steady-state correlation between i 's and j 's deviations from average is

$$C_{zij} = \frac{\Sigma_{ss,zij}}{\sqrt{\Sigma_{ss,zii} \Sigma_{ss,zjj}}}$$

To find the correlation length λ for a given set of strategies, we find the distances between all pairs of birds and $\{C_{zij}\}$ for 1000 cues. In a unit square, the maximum possible distance between a pair of birds is $\sqrt{2}$. We divide the range $[0, \sqrt{2}]$ into a number of small bins. For each bin, we find the average across all cues of the correlations between pairs of birds whose distance is within the bin. When we plot this average correlation against the center of each bin, the correlation length λ is the distance at which the correlations cross 0. (Such a distance must exist by Fact A.2.) If birds i and j are within λ of each other, then the fluctuations in the deviations of their opinions from the flock average will tend to be positively correlated. An illustration of this calculation is provided in Fig. 2.

The group-level properties we just defined, ρ and λ , rely on the steady-state covariance matrices of y and z respectively. To fully define these properties, we must therefore introduce noise into the system, which we do using the stochastic differential equations

$$dv = Mvdt + DdW_t,$$

where $M \in \mathbb{R}^{N \times N}$ is an interaction matrix, $D \in \mathbb{R}^{N \times N}$ is a diagonal matrix, and W is an N -dimensional Wiener process. Each property will require a different interaction matrix M and diagonal matrix D . However, regardless of our specific choice of M and D , as long as $\mathbf{1}$ is an eigenvector of M with eigenvalue 0 and all other eigenvalues of M have negative real part, we can find the covariance matrices of y and z as follows. Following Young et al. (2010, 2013), we use Ito's formula to find the stochastic differential equations that describe y :

$$\begin{aligned} dy &= QMvdt + QDdW_t \\ &= QM(z + \bar{v}\mathbf{1})dt + QDdW_t \\ &= QM(Q^T y + \bar{v}\mathbf{1})dt + QDdW_t \\ &= QMQ^T ydt + QDdW_t \text{ since } M\mathbf{1} = \mathbf{0}, \\ &= \bar{M}dt + \bar{D}dW_t \end{aligned}$$

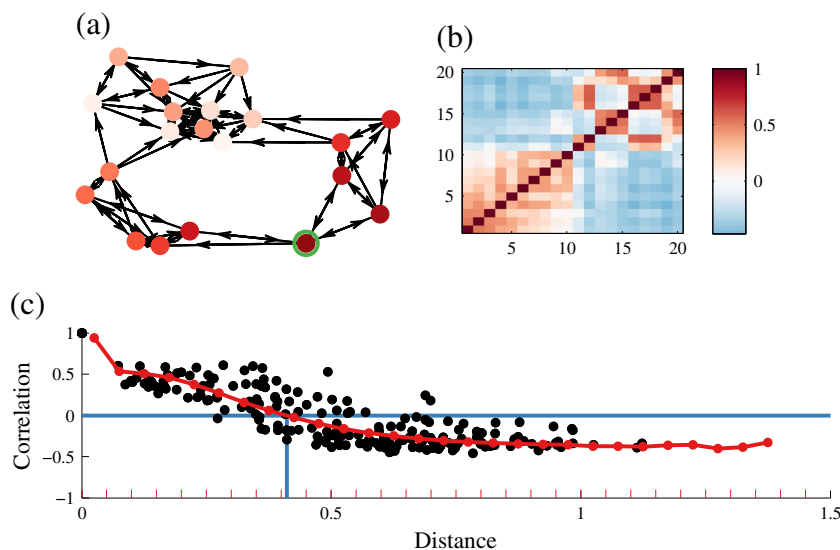


Fig. 2 An illustration of how to calculate correlation length. In **a**, we show a group of $N = 20$ birds, each of which has four neighbors. The bird circled in green perceives the cue. The color of each node indicates the opinion of the corresponding bird at $t = 1$. An arrow points from i to j if i pays attention to j . In **b**, we show the steady state correlation between each pair of birds in this network. In **c**, we show how we find the correlation length of a flock (in this case, one in which each bird has four neighbors). The black points show the correlation

between each pair of birds as a function of the distance between them, for the network shown in **(a)**. The red points and line show the average correlation of pairs of birds whose distance falls within the bins indicated on the x-axis, where the average is taken across 1000 cues. The horizontal blue line shows correlation equal to 0 and the vertical blue line shows the distance at which the correlation function crosses 0, i.e., the correlation length. Parameters: $N = 20, r = 0.2, b = 1$

where $\bar{M} = QMQ^T$ and $\bar{D} = QD$. Therefore, $y(t)$ is described by a Gaussian probability density, whose covariance matrix $\Sigma_y(t)$ obeys the differential equations (Arnold 1974)

$$\frac{d\Sigma_y}{dt} = \bar{M}\Sigma_y + \Sigma_y\bar{M} + \bar{D}\bar{D}^T.$$

As shown in Young et al. (2010, 2013), the $N - 1$ eigenvalues of \bar{M} are the same as the $N - 1$ non-zero eigenvalues of M , so all the eigenvalues of \bar{M} have negative real part. Therefore, the steady-state covariance matrix $\Sigma_{ss,y}$ satisfies the Lyapunov equation (Young et al. 2010, 2013)

$$\bar{M}\Sigma_{ss,y} + \Sigma_{ss,y}\bar{M} = -\bar{D}\bar{D}^T. \tag{2}$$

(Full statements of these claims are provided in the Appendix A.2. See also refs. Bamieh et al. 2012; Xiao et al. 2007; Zelazo and Mesbahi 2009.) Since $z = Q^T y$, we can find $\Sigma_{ss,z}$ using $\Sigma_{ss,z} = Q^T \Sigma_{ss,y} Q$. Although all the results we present about group-level properties are in the absence of an environmental cue, in Appendix A.3, we show how to find $\Sigma_{ss,y}$ when an environmental cue is present. (see also Fitch and Leonard 2014; Pais and Leonard 2014). To fully specify these matrices, it remains to specify M and D , which will depend on the group-level property we are calculating.

To calculate ρ , we use $M = L$, as in Young et al. (2013). By Gershgorin's circle theorem, all eigenvalues of L have zero or negative real part. By definition of L , $\mathbf{1}$ is an eigenvector

of L with eigenvalue 0 and, if the network is connected, there is only one 0 eigenvalue (Agaev and Chebotarev 2005; Young et al. 2010). (If the network is not connected, as may happen when the birds have few neighbors, we set $\rho = 0$.) We set the diagonal elements of D equal to $\{\sqrt{k_1}, \dots, \sqrt{k_n}\}$. This is equivalent to using $D = I$ and subsequently dividing ρ by \sqrt{k} in the case of a flock whose members all use the same strategy, as was done in Young et al. (2013). Because of this normalization, what we call ρ , Young et al. (2013) called "robustness per neighbor." With these choices of M and D , we use Eq. 2 to find the steady-state covariance matrix, $\Sigma_{ss,y}$.

In statistical mechanics, it is conventional to assume symmetric interactions between the components of a system, here the birds. Therefore, to calculate λ , we symmetrize A by defining $S = \frac{1}{2}(A + A^T)$ and

$$P_{ij} = \begin{cases} S_{ij} & \text{if } j \neq i \\ -\sum_l S_{il} & \text{if } j = i. \end{cases}$$

If A is symmetric, then $S = A$ and $P = L$. By definition of P , $\mathbf{1}$ is an eigenvector of P with eigenvalue 0 and, if the network is connected, all other eigenvalues have negative real part, according to theorems in Agaev and Chebotarev (2005) and Young et al. (2010). We can now use $M = P$ and $D = I$ in Eq. 2. After using Eq. 2 to find $\Sigma_{ss,y}$, we can calculate the covariance matrix $\Sigma_{ss,z} = Q^T \Sigma_{ss,y} Q$ and then the correlation matrix C_z . (If the network is not connected, we use the procedure described in Appendix A.4 to find C_z . In this

case, the correlation between birds in different components of the network will be 0.)

There is one last group-level property that will be useful in understanding the social context that determines whether a strategy is more or less advantageous. For this measure, we return to the ordinary opinion dynamics when there is no cue present,

$$\frac{dv}{dt} = Lv.$$

If σ denotes the absolute value of the non-zero eigenvalue of L closest to 0, then σ is the speed with which the flock converges to consensus (Olfati-Saber and Murray 2004; Young et al. 2010).

To find the correlation length of starling flocks, Bialek et al. (2014) wrote down a Boltzmann distribution over opinion vectors, rather than considering stochastic dynamics leading to a steady-state distribution. In Appendix A.4, we show how to find the covariance matrix using this approach and in Appendix A.5, we show that the covariance matrix found using stochastic dynamics and the Lyapunov equation (2) is equivalent to that found using the Boltzmann distribution, as long as the same M and D are used in both approaches. Figure 7 shows that even if $M = L$ is used in the Lyapunov equation and $M = P$ is used in the Boltzmann distribution, the resulting covariance matrices $\Sigma_{ss,z}$ are almost identical.

Results

The ESS number of neighbors depends on the content of the cue

The ESS number of neighbors depends on which measure of fitness we use. The ESS number of neighbors when fitness

is based on access to scarce resources or on both access to scarce resources and predation is always low, while the ESS number of neighbors when fitness is based on predation is always high (Fig. 3). There is no ESS when fitness is based on access to abundant resources.

According to the approximation in Eq. 1, a bird is better informed if a higher proportion of its neighbors perceive the cue. To avoid being eaten, a bird must therefore avoid having a low proportion of its neighbors perceive the cue. To get access to scarce resources, on the other hand, a bird must have a high proportion of its neighbors perceive the cue as often as possible. To show why different strategies are needed to attain these different objectives, we consider a simpler situation in which each bird chooses k_i neighbors at random from the other $N - 1$ birds, instead of choosing the closest birds. If N_c birds perceive the cue and i is not one of these, then the number of i 's neighbors that do perceive the cue follows a hypergeometric distribution:

$$P(j \text{ of } i\text{'s neighbors perceive the cue}) = \frac{\binom{N_c}{j} \binom{N-1-N_c}{k_i-j}}{\binom{N-1}{k_i}}.$$

If p_i is the proportion of i 's neighbors that perceive the cue, it immediately follows that

$$P\left(p_i = \frac{j}{k_i}\right) = \frac{\binom{N_c}{j} \binom{N-1-N_c}{k_i-j}}{\binom{N-1}{k_i}}. \tag{3}$$

In the Appendix A.2 and Fig. 8, we show that the mean value of p_i does not depend on k_i and that the variance of p_i decreases as a function of k_i . Therefore, if two birds use strategies k_1 and k_2 with $k_1 < k_2$, then their opinions at

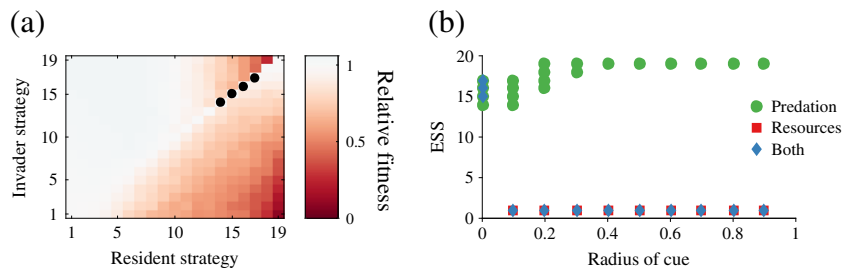


Fig. 3 The ESS number of neighbors is always higher if fitness is based on predation than if fitness is based on scarce resources or on both. When selection is due to predation, the (possibly multiple) ESS strategies increase with the radius of the cue. In **a**, we show the relative fitness of an invader as a function of the invader and resident strategies, where fitness is based on predation and the radius $r = 0$. The points indicate the ESS strategies. In **b**, we show the (possibly multiple) ESS

strategies as a function of the radius of the cue. The green circles show the ESS when fitness is based on predation, the red squares when fitness is based on access to scarce resources, and the blue diamonds when fitness is based on both. The squares and diamonds overlap when the radius is greater than 0. There are no ESS when fitness is based on access to abundant resources. Parameters: $N = 20, b = 1$

the end of many cues will be the same on average, but the variance of the opinions of the bird using k_1 will be higher. This explains why a high strategy minimizes the probability of having a poorly-informed opinion and a low strategy maximizes the probability of having a well-informed opinion, leading to the ESS number of neighbors we find under different measures of fitness.

The multiple predation ESS increase as the cue becomes more widespread

The ESS numbers of neighbors when fitness is based on avoiding predation exhibit two interesting features: there are sometimes multiple consecutive ESS and the ESS number of neighbors increases as the cue becomes more widespread (Fig. 3). It is somewhat surprising to find, when $r = 0$ for example, that the strategy of 14 neighbors is uninvadable by the strategy of 15 neighbors and also that the strategy of 15 neighbors is uninvadable by the strategy of 14 neighbors. This is because there is positive frequency dependent selection. When $r = 0$, there is only one bird in the flock who perceives the environmental cue. Consider a bird i who has 14 neighbors and a bird j who has 15 neighbors. Bird j will be more likely to pay attention to the bird that perceives the cue, which imposes a cost on i . However, when they both happen to pay attention to the bird with the cue, then the proportion of j 's neighbors with the cue is lower than the proportion of i 's neighbors with the cue, which imposes a

cost on j . The relative magnitude of these costs depends on the strategies the rest of the flock is using.

A flock all of whose birds use the same strategy k converges to $\mathbf{1}$ more quickly as k increases (Fig. 9). In other words, if most birds in a flock have 14 neighbors, they will be generally worse-informed than the birds in a flock where most of the birds have 15 neighbors. Therefore, it is more costly for j to pay attention to birds without the cue in a flock most of whose birds have 14 neighbors than in a flock most of whose birds have 15 neighbors (Fig. 4a). Conversely, if i does not pay attention to the bird with the cue, i 's slow learning rate is much worse than its peers' learning rates in a flock where most of the birds have 15 neighbors, whereas it performs just as well as its peers in a flock where most of the birds have 14 neighbors (Fig. 4a). Therefore, whichever strategy the minority of the flock is using cannot invade, leading to multiple ESS.

To understand why the ESS number of neighbors increases as the cue becomes more widespread, again consider two birds, i with 18 neighbors and j with 19 neighbors. When the radius of the cue is low and there are few birds that perceive the cue, i is no less likely to pay attention to the few perceiving birds and i s more likely to have a higher proportion of its neighbors perceive the cue, so 18 can invade 19 (Fig. 4b). As the radius increases and more birds perceive the cue, i is now more likely to pay attention to fewer of the perceiving birds, so that 18 can no longer invade 19 and 19 becomes an ESS (Fig. 4c).

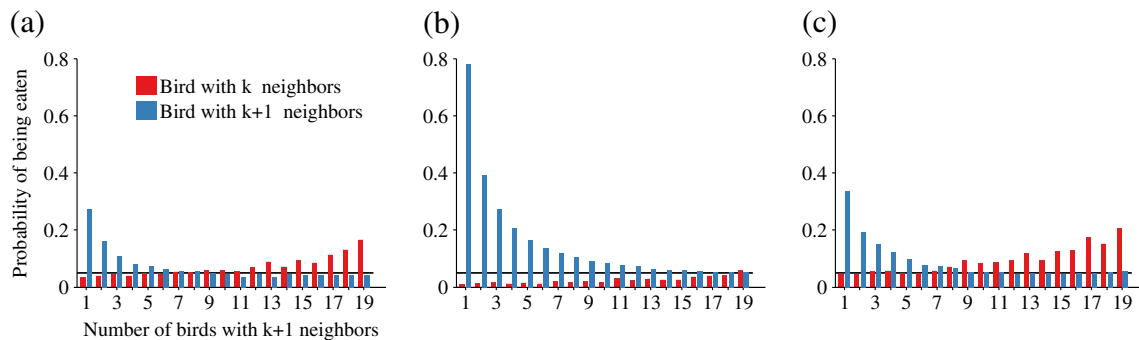


Fig. 4 The fitness, due to predation, of a strategy can increase as it becomes more frequent in the flock. In each panel, we consider a flock made up of 20 birds, each of which uses one of two strategies, k or $k + 1$. The number of birds in the flock with $k + 1$ neighbors is shown on the x-axis. We then focus on two birds in the flock, i using k neighbors and j using $k + 1$ neighbors. In each pair of bars, the red bar on the left corresponds to i and the blue bar on the right corresponds to j . The height of the bar indicates the probability that the bird is eaten; the black line shows probability $0.05 = 1/20$. In a, $k = 14$, $k + 1 = 15$, and $r = 0$ (one bird perceives the cue). When there are only a few birds using 15 neighbors, j has a high probability of being eaten because it is likely that too low a proportion of its neighbors perceive the cue, in comparison to the rest of the flock. When there are many birds using 15 neighbors, i has a high probability of being eaten because it is more likely than the rest of the flock not to pay attention to the perceiver. This results in neither strategy being able to invade the

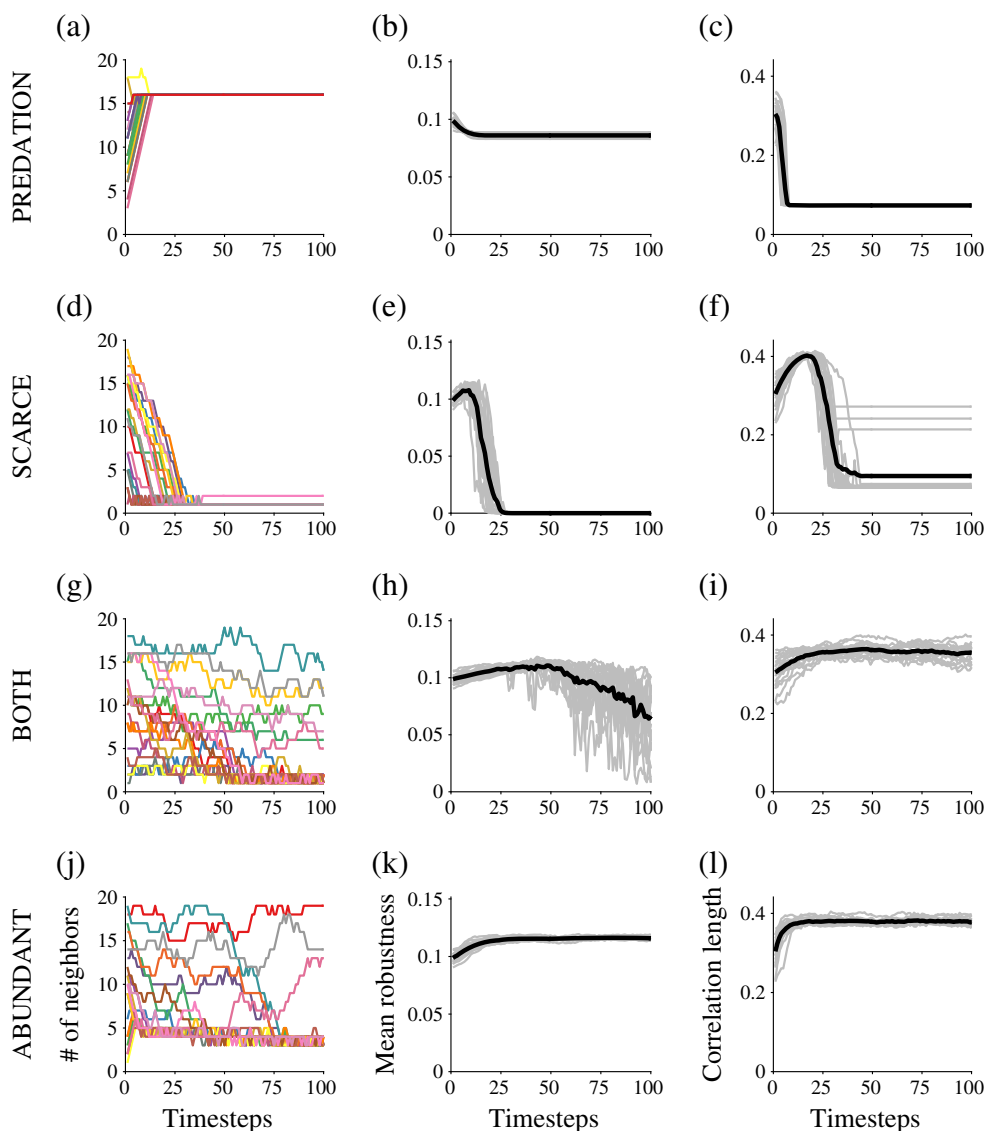
other. In b, $k = 18$, $k + 1 = 19$, $r = 0$ (one bird perceives the cue). When there are only a few birds using 19 neighbors, j has a high probability of being eaten because it is likely that too low a proportion of its neighbors perceive the cue, in comparison to the rest of the flock. However, regardless of the composition of the flock, i is likely to pay attention to the perceiver and never has a probability of being eaten that is greater than chance. This results in 18 being able to invade 19, but not conversely. In c, $k = 18$, $k + 1 = 19$, $r = 0.3$. When there are only a few birds using 19 neighbors, j has a high probability of being eaten because it is likely that too low a proportion of its neighbors perceive the cue, in comparison to the rest of the flock. When there are many birds using 19 neighbors, i has a high probability of being eaten because it is likely to pay attention to fewer of the perceivers than the rest of the flock. This results in neither strategy being able to invade the other. Parameters: $N = 20$, $b = 1$

The birds construct an optimal network only under certain selection regimes

When every bird is using the same strategy k , the average robustness $\bar{\rho}$ and the correlation length λ are both maximized at intermediate strategies and minimized at extreme high or low strategies (Fig. 10). The greedy optimization process allows us to analyze how the group-level properties of a heterogeneous flock change as the birds choose new strategies. The strategies that result from greedy optimization are similar to the ESS under different selection regimes. When fitness is based on avoiding predation, the birds choose high strategies, ultimately converging on one of the ESS for the same parameters (Fig. 5). The convergence to a single strategy can be explained by the positive frequency dependence of this type of selection: once a

strategy has a slight majority, all birds have an incentive to adopt it. This leads to moderately high robustness, but low correlation length. When fitness is based on access to scarce resources, the birds all choose low strategies. This leads to both low robustness and low correlation length. When fitness is based on both access to scarce resources and predation, a few birds choose low strategies, but the flock remains quite heterogeneous. The heterogeneity in the flock keeps the correlation length high, while the robustness still tends to decrease. When fitness is based on access to abundant resources, most birds choose strategies of three to five neighbors, but a few choose high strategies. This is the selection regime that leads the birds to construct a flock with both high robustness and high correlation length. We find similar results when abundant resources means that either two or ten birds are allowed to eat, rather than four (Fig. 11).

Fig. 5 When the birds choose the strategies that improve their own fitness, they settle on strategies close to the ESS. Only fitness based on access to abundant resources leads to both high robustness and high correlation length. The *upper row* shows results for fitness based on predation, the *second row* shows results for fitness based on access to scarce resources, the *third row* shows results for fitness based on both scarce resources and predation, and the *last row* shows results for fitness based on access to abundant resources. The *first column* shows one example of how the birds' strategies change over time. The *second column* shows, for various initial conditions, how the average robustness, $\bar{\rho}$, changes over time, with the *black line* showing the average of the 20 *gray lines*. The *third column* shows, for various initial conditions, how the correlation length, λ , changes over time, with the *black line* showing the average of the 20 *gray lines*. Parameters: $N = 20, b = 1, r = 0.1$



Discussion

Here, we study a simple theoretical model of opinion dynamics driven by environmental cues with the goals of (1) understanding how the content of the cue affects the evolution of social-information gathering strategies and (2) testing whether evolution acting on individuals' strategies can lead to optimal group-level properties. We find that the content of the environmental cue strongly affects what number of neighbors a bird should use to gather social information, with a cue about predators leading to high numbers of neighbors and a cue about resources leading to low numbers of neighbors. We also find that as information about predators becomes more widespread, it is advantageous to pay attention to more neighbors. Only when the cue is about abundant resources does a greedy optimization process lead to the flock using strategies that optimize both the robustness with which it can reach consensus and the correlation length of the flock.

While there is a growing body of work on how statistical properties of information, like its variance, affect the evolution of information-gathering strategies (e.g., Henrich and Boyd 1998; King and Cowlshaw 2007; Shaw and Couzin 2013), less work has been done to study the effects of the content of that information. Our results show that whether a cue is about predators or resources can have a dramatic effect on the optimal way of gathering social information. We find the same effect—high strategies for cues about predators and low strategies for cues about resources—when we analyze both the ESS numbers of neighbors and the numbers of neighbors to which a greedy optimization process converges. This suggests that our results are robust to changes in how these strategies are acquired, for example if they are genetically inherited or learned. In our model, there are no costs to using more neighbors. If there were, we might expect the high ESS number of neighbors under predation to decrease. However, without costs to higher strategies, our model allows us to analyze the advantages and disadvantages of the information-gathering strategies themselves. While the model was inspired by studies of starling flocks, it should be general enough to apply to any species in which topological rules determine how animals gather social information. Our results lead to an interesting prediction that species whose fitness depends more strongly on avoiding predation than on having access to resources might rely more heavily on social information. Testing this prediction by studying the social interactions in multiple species will be an avenue for future research.

There are two ways in which we might expect that the reliability and availability of personal information to affect the use of social information. On the one hand, it might only be worthwhile to gather information from other individuals

if they have good personal information about the environment and if there are many individuals who have such information. This is what we find in our model when the cue is about predators: as more birds in the flock receive information, it becomes advantageous to pay attention to more neighbors. Our result is similar to the findings of King and Cowlshaw (2007), who used a model of majority decision-making to conclude that individuals did better by pooling their information only when their personal information was quite reliable. Increasing the number of animals with personal information also leads to an increase in the payoff from using social information in producer-scrouter games (Giraldeau and Dubois 2008; Rieucou and Giraldeau 2011; Vickery et al. 1991). As a final example of this type of reasoning, Giraldeau et al. (2002) argued that using social information can result in disadvantageous informational cascades when environmental cues are not entirely reliable and animals only have information about their peers' response to the environment (rather than about the cues their peers perceive); they further suggest that the use of social information may be less likely to evolve in these cases. On the other hand, individuals might be forced to rely on social information more when they do not have access to reliable personal information. There are multiple studies whose authors use some variant of this argument. Henrich and Boyd (1998) used a quite general theoretical model to conclude that we should expect social-information use to evolve when an individual's perceptions of the environment are unreliable. Similarly, Shaw and Couzin (2013) studied a model describing animals that could rely on historical information, information about current resources, and social information in order to migrate successfully and found that animals should evolve higher dependence on social information when the environment changes more from year to year. Whether the reliability and availability of personal information incentivizes or disincentivizes the use of social information seems to be quite model-dependent, and in order to make predictions about any given empirical system, we would have to identify carefully which model best described it.

Finding the ESS number of neighbors led us to identify an interesting positive frequency-dependent selection. Specifically, when the environmental cue is about predators, the strategy being used by the majority sets a social context that can prevent other strategies from invading. Conversely, Torney et al. (2015) found that an individual using a high amount of social information can do well when that is the minority strategy. Frequency dependence seems to be a common property of systems in which animals use social information (Rieucou and Giraldeau 2011). These findings highlight the complexities that can arise in studying the evolutionary dynamics of social behaviors when the fitness of a

strategy depends on the social environment of the individual using it.

The fact that the costs and benefits of using a particular strategy to gather social information depends on the social context means there is no single best strategy, so heterogeneity in the population can persist. We saw this when we looked at greedy optimization processes for fitness based on access to abundant resources or based on both predation and access to scarce resources. Pais and Leonard (2014) also found that heterogeneity in the weight given to social information can persist, both when the trait is determined by adaptive dynamics and when individuals selfishly change their strategies to improve their own fitness. That different models of different types of strategies give rise to heterogeneity suggests that it might be a common feature of strategies for gathering social information and that we might expect to see it in natural populations.

Since in our model the content of the environmental cue affects what strategies should be used to gather social information, it also affects whether or not the birds construct an optimal network. We find that greedy optimization leads to a flock having either low robustness or low correlation length or both, except when the birds choose strategies that maximize their ability to get access to abundant resources. This agrees with previous findings that selection acting on individuals' strategies can lead to situations that are suboptimal from the group's perspective (Torney et al. 2015). Our result shows that the strategies that optimize robustness and correlation length do not necessarily optimize a bird's ability to learn about its environment. However, there may be other advantages to being in a group that can reach consensus that provide incentives for birds to construct networks with high robustness. In addition to the number of neighbors being used, which birds attend to the cue affects how well the flock can reach consensus and there are well-defined sets of "leaders" that maximize the flock's robustness (Fitch and Leonard 2014). It would be interesting to study whether there is an evolutionary process by which these individuals could be incentivized to take on that role.

Our model is not directly comparable to the empirical starling flocks to which these measures have been applied because we consider many fewer birds than are found in most real groups and we use random positions rather than empirically-derived positions. Additionally, since we only considered one group size ($N = 20$), we could not reproduce the linear scaling of correlation length with group size that has been used to argue that starling flocks are near criticality (Bialek et al. 2014; Cavagna et al. 2009). However, our observation that robustness first increases and then decreases with the number of neighbors agrees qualitatively with empirical observations (Young et al. 2013), suggesting a coarse agreement between our findings and real starling

flocks. In both our model and the empirical measurements, this non-monotonic behavior occurs because, even though adding more neighbors tends to make the interaction network more dense and improve how well the flock can reach consensus, eventually these improvements are outweighed by the decreases caused by dividing by the square root of k . Another important difference between our model and real starling flocks is that we only consider two-dimensional space. In real flocks, robustness decreases as the flock becomes "flatter" and closer to two-dimensional (Young et al. 2013), so the number of dimensions the birds have access to matters. However, we expect that the general patterns we observe would hold in a three-dimensional model.

We intentionally built a simple model in order to analyze the consequences of a small number of assumptions and parameters. Bialek et al. (2014) modeled the opinions of birds about where to fly using a three-dimensional opinion vector, but Young et al. (2013) used a scalar opinion, a simplifying assumption we also used. We expect that our results would be similar were we to use three-dimensional opinions. While the question of whether real birds use metric or topological interaction neighborhoods is still unresolved, we chose to use topological neighborhoods in order to make our results comparable to the empirical observations of Bialek et al. (2014) and Young et al. (2013). Rather than finding the evolutionary equilibrium number of neighbors, we could find the evolutionary equilibrium range of interaction, and we expect that our results about the effects of the content of information and the response to how widespread the cue is should be similar in such a model. While we allow the birds to move between cues, a major simplifying assumption is that the interaction network is static while the cue is present. This is a reasonable assumption if the fitness consequences of the cue occur so quickly after the cue appears that the birds do not have time to move enough to change their interaction network. Even if the interaction network were changing as the birds respond to the cue, we would expect similar results since the numbers of neighbors should similarly affect each new interaction network. Coupling our model of opinion dynamics and cues with rules for how the birds move through space would be an interesting extension.

By analyzing a simple model of opinion dynamics in flocking birds, we make novel predictions about how the content of information affects the evolutionary equilibria of strategies for its use and we find that these strategies would only lead to optimal group-level properties when the information was about certain facets of the environment. This makes the observations that starling flocks have high robustness and are near criticality all the more remarkable. Our results also provide insight into how to incentivize

individuals to interact with each other in such a way that they reach a stable consensus, if such an outcome is desirable.

Acknowledgments ERB acknowledges support from the training grant NIH 5T32HG003284, from a subaward from the Santa Fe Institute under a grant from the John Templeton Foundation for the study of complexity, and from Princeton University through the Charlotte Elizabeth Procter Fellowship. NEL acknowledges support from ONR grant N00014-14-1 0635. SAL acknowledges support from NSF grant GEO-1211972 and NSF grant OCE-1426746. NEL and SAL acknowledge support from ARO grant W911NF-14-1-0431 and ARO grant W911NF-11-1-0385. The preparation of this paper has been supported by the Nordforsk-funded project Green Growth Based on Marine Resources: Ecological and Socio-Economic Constraints (GreenMAR).

Appendix A

A.1 ESS in finite populations

In this section, we explain how to find the evolutionarily stable strategy in a finite population. In an infinite population, a strategy is evolutionarily stable if the fitness of an individual using any other initially rare strategy is lower than the fitness of an individual using the focal strategy (Nowak 2006). In other words, a strategy is an ESS if, once all members of a population use it, no other strategy can invade. In a population of finite size N , stochastic events could lead to a strategy spreading through the population (reaching fixation) or going extinct, despite differences in individuals' fitnesses Nowak (2006). Therefore, a little more care is needed to define and find an ESS in finite populations.

To define an ESS in a finite population Nowak (2006) considers a Moran process on a finite population: at each timestep, an individual is chosen to die at random and an individual is chosen to reproduce with a probability proportional to its fitness. In a population of size N , if there are $N - n$ "resident" individuals using strategy k and n "mutant" individuals using strategy k' , we will write f_n for the fitness of the mutants and g_n for the fitness of the residents, following Nowak (2006). In the Moran process, the probability of increasing from n to $n + 1$ individuals using k' is

$$p_{n,n+1} = \frac{nf_n}{nf_n + (N - n)g_n} \frac{N - n}{N},$$

the probability of decreasing from n to $n - 1$ individuals using k' is

$$p_{n,n-1} = \frac{(N - n)g_n}{nf_n + (N - n)g_n} \frac{n}{N},$$

and probability of the number of individuals using k' remaining at n is

$$p_{n,n} = 1 - p_{n,n+1} - p_{n,n-1}.$$

Using these transition probabilities, Nowak (2006) then derives the probability of k' reaching fixation when the population starts with one individual using k' and the rest using k :

$$p_{k',k} = \frac{1}{1 + \sum_{\ell=1}^{N-1} \prod_{n=1}^{\ell} \frac{g_n}{f_n}}. \tag{4}$$

If the two strategies have equal fitness, so that $f_n = g_n$ for all n , then the probability of k' reaching fixation is $1/N$. This gives us a neutral baseline to which we can compare the probability of a rare strategy reaching fixation.

As described above, a strategy is an ESS in an infinite population if no other strategy can invade when initially rare, which would happen if the fitness of the invader were greater than the fitness of the resident. We can now supplement this condition to define an ESS in a finite population. Specifically, according to Nowak (2006), a strategy k is an ESS if, for all other strategies k' , both of the following conditions hold:

1. Selection opposes k' invading $k \iff f_1 < g_1$ and
2. Selection opposes k' replacing $k \iff p_{k',k} < 1/N$.

In our model, the strategies we consider are the number of neighbors to which a bird pays attention, which can range from 1 to $N - 1$. To find the ESS strategy, for each pair of strategies, invader k' and resident k , and for each $n = 1, \dots, N - 1$, we consider a flock in which n individuals use strategy k' and $N - n$ individuals use strategy k . We find the fitness for each individual using the procedure described in the main text. We then define f_n as the average fitness of the individuals using k' and g_n as the average fitness of the individuals using k . Equation 4 allows us to calculate the probability of strategy k' reaching fixation in a population using k , $p_{k',k}$. Using f_1 , g_1 , and $p_{k',k}$, we can finally check the conditions for k being an ESS strategy. In Fig. 6, we illustrate how we find the ESS number of neighbors when fitness is based on predation.

A.2 Mathematical background

In this section, we give the full statements of claims that are used in the text and other sections of the Appendix, and we provide proofs that are original to this paper or references to proofs of widely known theorems.

Claim 1 If $v \in \mathbb{C}^N$ is an eigenvector of L_s with eigenvalue λ , then

$$\frac{\sum_{i \text{ perceivers}} |v_i|^2}{\sum_{i=1}^N |v_i|^2} \geq \frac{|\lambda| - 2}{b}.$$

Proof By Gershgorin's circle theorem, if $\rho(L) = \max_i \{|\lambda_i| \text{ s.t. } \lambda_i \text{ is an eigenvalue of } L\}$ is the spectral radius of

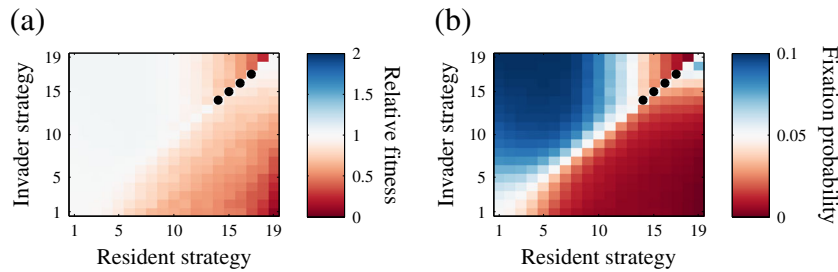


Fig. 6 A strategy is evolutionarily stable if selection opposes any other strategy invading it and selection opposes any other strategy replacing it. A mutant strategy k' can invade the resident strategy k if the fitness of the only individual in the population using it, f_1 , is greater than the fitness of the individuals using the resident strategy, g_1 . A mutant strategy k' can replace the resident strategy k if the probability of k' reaching fixation is greater than chance, $1/N = 0.05$.

L , then $\rho(L) \leq 2$. If $v \in \mathbb{C}^N$ is an eigenvector of L_s with eigenvalue λ ,

$$\begin{aligned}
 v^* L_s v &= \lambda v^* v, \text{ where } v^* \text{ is the conjugate} \\
 &\quad \text{transpose of } v \\
 \Rightarrow |v^* L_s v| &= |\lambda| \cdot |v^* v| \\
 \Rightarrow |v^* L v - v^* B v| &\geq |\lambda| \cdot |v^* v| \\
 \Rightarrow |v^* L v| + |v^* B v| &\geq |\lambda| \cdot |v^* v| \text{ by the triangle} \\
 &\quad \text{inequality} \\
 \Rightarrow \rho(L) |v^* v| + |v^* B v| &\geq |\lambda| \cdot |v^* v| \\
 \Rightarrow 2|v^* v| + |v^* B v| &\geq |\lambda| \cdot |v^* v| \\
 \Rightarrow b \sum_{i \text{ perceivers}} \bar{v}_i v_i &\geq (|\lambda| - 2) \sum_i \bar{v}_i v_i, \text{ where } \bar{v}_i \text{ is} \\
 &\quad \text{the complex conjugate of } v_i \\
 \Rightarrow \frac{\sum_{i \text{ perceivers}} |v_i|^2}{\sum_{i=1}^N |v_i|^2} &\geq \frac{|\lambda| - 2}{b}
 \end{aligned}$$

□

Claim 2 Suppose $v(t) \in \mathbb{R}^N$ satisfies the differential equations $\dot{v} = Xv + c$, where $X \in \mathbb{R}^{N \times N}$ and $c \in \mathbb{R}^N$. Suppose all of the eigenvalues of X are non-zero and have small absolute value. If V is the matrix whose columns are eigenvectors of X and Λ is the matrix with the eigenvalues of X along the diagonal, we can approximate $v(t) \approx v_0 + V(tI + \frac{t^2}{2}\Lambda)V^{-1}c$. In the case that $X = L_s$ and $c = B\mathbf{1}$, for i that do not perceive the cue,

$$v_i(t) \approx v_i(0) + \frac{t^2 b}{2} \sum_{\text{perceivers } j} L_{ij}.$$

Proof We first write $v(t) = \exp(tX)v_0 + \exp(tX) \int_0^t \exp(-sX)cds$. If X has no 0 eigenvalues,

$$v(t) = V \exp(t\Lambda)V^{-1}v_0 + V\Lambda^{-1}(\exp(t\Lambda) - I)V^{-1}c$$

Here, we consider the strategy of the number of neighbors to have, where fitness depends on the probability of surviving predation. In **a**, we show the relative fitness, f_1/g_1 . In **b**, we show the probability of reaching fixation, $p_{k',k}$. The four dots in each panel indicate the four ESS. Parameters: $N = 20, b = 1, r = 0$

If $h(x) = \exp(tx)$, then $h(x) \approx 1$ for x close to 0. If $g(x) = \frac{1}{x}(\exp(tx) - 1)$, then $g(x) \approx t + \frac{t^2}{2}x$ for x close to 0. Thus, if the eigenvalues of X are close to 0, we find that

$$\begin{aligned}
 v(t) &\approx VIV^{-1}v_0 + V\left(tI + \frac{t^2}{2}\Lambda\right)V^{-1}c \\
 &= v_0 + tc + \frac{t^2}{2}Xc.
 \end{aligned}$$

If we use $X = L_s$ and $c = B\mathbf{1}$, we find

$$v(t) \approx v_0 + tB\mathbf{1} + \frac{t^2}{2}L_s B\mathbf{1}.$$

For birds that do not perceive the cue, this gives

$$v_i(t) \approx v_i(0) + \frac{t^2 b}{2} \sum_{\text{perceivers } j} L_{ij}.$$

By Claim A.2, if $v \in \mathbb{C}^N$ is an eigenvector of L_s with eigenvalue λ that has a large absolute value, then the elements of v corresponding to birds that do not perceive the cue will be small. Therefore, even if there are eigenvalues of L_s far from 0, our approximation holds for elements of v corresponding to birds that do not perceive the cue. □

Claim 3 If there are $N - 1$ objects, exactly N_c of which are “successes”, if k objects are drawn at random without replacement, and if p is the proportion of draws that are successes, then for $j = 0, \dots, k$,

$$P\left(p = \frac{j}{k}\right) = \frac{\binom{N_c}{j} \binom{N-1-N_c}{k-j}}{\binom{N-1}{k}}.$$

The expectation of p is $E[p] = \frac{N_c}{N-1}$ and the variance of p is

$$\text{Var}(p) = \frac{(N-1-k)N_c(N-1-N_c)}{k(N-1)^2(N-2)}. \tag{5}$$

$\text{Var}(p)$ decreases as a function of k and is minimized at $k = N - 1$, where $\text{Var}(p) = 0$.

Proof The number of successes j follows a hypergeometric distribution:

$$P(j \text{ successes}) = \frac{\binom{N_c}{j} \binom{N-1-N_c}{k-j}}{\binom{N-1}{k}}.$$

Since $P\left(p = \frac{j}{k}\right) = P(j \text{ successes})$, it is clear that the probability distribution of p is

$$P\left(p = \frac{j}{k}\right) = \frac{\binom{N_c}{j} \binom{N-1-N_c}{k-j}}{\binom{N-1}{k}}.$$

The expectation of the hypergeometric distribution is $k \frac{N_c}{N-1}$, so the expectation of p is $E[p] = \frac{N_c}{N-1}$. The variance of the hypergeometric distribution is

$$\begin{aligned} \text{Var}(\# \text{ of successes}) &= (N-1-k)k \times \frac{N_c(N-1-N_c)}{(N-1)^2(N-2)} \\ \Rightarrow \text{Var}(p) &= \frac{1}{k^2} \text{Var}(\# \text{ of successes}) = \frac{N-1-k}{k} \frac{N_c(N-1-N_c)}{(N-1)^2(N-2)} \end{aligned}$$

We are considering a flock of size N , N_c of which perceive the cue, and at least 1 of which does not. If $N = 2$, then for i that does not perceive the cue, p is always 1 and $\text{Var}(p) = 0$. If $N > 2$, then $\frac{N_c(N-1-N_c)}{(N-1)^2(N-2)} \geq 0$ since $N-1 \geq N_c \geq 1$. Let $f(k) = \frac{N-1-k}{k}$. Then

$$\begin{aligned} f(k+1) - f(k) &= \frac{N-1-k-1}{k+1} - \frac{N-1-k}{k} \\ &= \frac{(N-1-k-1)k - (N-1-k)(k+1)}{(k+1)k} \\ &= \frac{-k - (N-1-k)}{(k+1)k} \\ &= \frac{-N+1}{(k+1)k} \\ &< 0 \text{ since } k \geq 1 \text{ and } N > 1 \end{aligned}$$

This shows that the variance of p decreases as a function of k and that it is therefore minimized at $k = N - 1$, at which $\text{Var}(p) = \frac{(N-1-(N-1))}{N-1} \frac{N_c(N-1-N_c)}{(N-1)^2(N-2)} = 0$. In Fig. 8, we show the distribution and variance of p for various values of k and N_c . \square

Theorem A.1 Given any positive symmetric matrix $Q > 0$, the following statements are equivalent:

1. There exists a unique positive symmetric matrix Σ that satisfies the Lyapunov equation $X\Sigma + \Sigma X^T = -Q$.

2. The system $\dot{y} = Xy$ is globally asymptotically stable.
3. All eigenvalues of X have negative real part.

Proof Proof available elsewhere (for example in Proposition 4.2 of Dullerud and Paganini 2000). \square

Theorem A.2 Consider the stochastic dynamical system,

$$dy = (Xy + c)dt + YdW_t$$

where $X \in \mathbb{R}^{N \times N}$ is such that all eigenvalues of A have negative real part, $c \in \mathbb{R}^N$, $Y \in \mathbb{R}^{N \times N}$, and W is an N -dimensional Wiener process. If we define the expectation vector $\mu(t) = E[y(t)]$, the covariance matrix $\Sigma(t) = E[(y(t) - \mu(t))(y(t) - \mu(t))^T]$, and the steady state covariance matrix $\Sigma_{ss} = \lim_{t \rightarrow \infty} \Sigma(t)$, then Σ_{ss} satisfies the Lyapunov equation

$$X\Sigma_{ss} + \Sigma_{ss}X^T = -YY^T. \tag{6}$$

The solution y_t is Gaussian if and only if y_0 is normally distributed or constant. In this case, at steady state,

$$\lim_{t \rightarrow \infty} y_t \sim \mathcal{N}(-X^{-1}c, \Sigma_{ss})$$

where Σ_{ss} is the solution to Eq. 6.

Proof This is a combination of Theorems 8.2.6, 8.2.10, and 8.2.12 in Arnold (1974). Note that by Theorem 8.2.6 in Arnold (1974), we can write the following matrix differential equation for $\Sigma(t)$:

$$\dot{\Sigma} = X\Sigma + \Sigma X^T + YY^T.$$

In Theorem 8.2.10, Arnold (1974) gives the solution

$$\Sigma_{ss} = \int_0^\infty \exp(Xs)YY^T \exp(X^T s) ds.$$

If we plug this solution into Eq. 6,

$$\begin{aligned} X\Sigma_{ss} + \Sigma_{ss}X^T &= \int_0^\infty \left(X \exp(Xs)YY^T \exp(X^T s) \right. \\ &\quad \left. + \exp(Xs)YY^T \exp(X^T s)X^T \right) ds \\ &= \int_0^\infty \frac{d}{ds} \left(\exp(Xs)YY^T \exp(X^T s) \right) ds \\ &= \left(\lim_{s \rightarrow \infty} \exp(Xs)YY^T \exp(X^T s) \right) - YY^T \\ &= -YY^T \text{ only if all the eigenvalues of} \\ &\quad \text{values of } X \text{ have negative real part.} \end{aligned}$$

Otherwise, the limit does not exist. In addition to proving that Σ_{ss} satisfies the Lyapunov equation, this provides an explicit formula for Σ_{ss} . \square

Claim A.2 If $\bar{v} = \sum_i v_i/N$ and $\Sigma = E[(v - \bar{v}\mathbf{1})^T(v - \bar{v}\mathbf{1})^T]$, then $\sum_{ij} \Sigma_{ij} = 0$.

Proof

$$\begin{aligned} \sum_{ij} E[(v_i - \bar{v})(v_j - \bar{v})] &= \sum_{ij} (E[v_i v_j] - E[v_i \bar{v}] \\ &\quad - E[v_j \bar{v}] + E[\bar{v} \bar{v}]) \\ &= \sum_{ij} E[v_i v_j] - N \sum_i E[v_i \bar{v}] \\ &\quad - N \sum_j E[v_j \bar{v}] + N^2 E[\bar{v} \bar{v}] \\ &= \sum_{ij} E[v_i v_j] - \sum_{ij} E[v_i v_j] \\ &\quad - \sum_{ij} E[v_i v_j] + \sum_{ij} E[v_i v_j] \\ &= 0. \end{aligned}$$

□

Claim A.2 Let $M \in \mathbb{R}^{N \times N}$ and $\tilde{M} \in \mathbb{R}^{(N-1) \times (N-1)}$ be the matrix derived from M by removing its last row and column. If we write the inverse of M as

$$M^{-1} = \begin{pmatrix} A & b \\ c^T & d \end{pmatrix}$$

where $b, c, \in \mathbb{R}^{N-1}$ and $d \in \mathbb{R}$, then

$$\tilde{M}^{-1} = A - \frac{1}{d}bc^T.$$

Proof We start with the definition of M^{-1} , following the exchange at <http://math.stackexchange.com/q/208021> (version: 2013-12-18):

$$\begin{aligned} MM^{-1} &= I_N \\ \Rightarrow \begin{pmatrix} \tilde{M} & \cdot \\ \cdot & \cdot \end{pmatrix} \begin{pmatrix} A & b \\ c^T & d \end{pmatrix} &= I_N \\ \Rightarrow \begin{pmatrix} \tilde{M} & \cdot \\ \cdot & \cdot \end{pmatrix} \begin{pmatrix} A & b \\ c^T & d \end{pmatrix} \begin{pmatrix} I_{N-1} & \mathbf{0} \\ -\frac{1}{d}c^T & 1 \end{pmatrix} &= \begin{pmatrix} I_{N-1} & \mathbf{0} \\ -\frac{1}{d}c^T & 1 \end{pmatrix} \\ \Rightarrow \begin{pmatrix} \tilde{M} & \cdot \\ \cdot & \cdot \end{pmatrix} \begin{pmatrix} A - \frac{1}{d}bc^T & b \\ c^T & d \end{pmatrix} &= \begin{pmatrix} I_{N-1} & \mathbf{0} \\ -\frac{1}{d}c^T & 1 \end{pmatrix} \\ \Rightarrow \begin{pmatrix} \tilde{M} & \cdot \\ \cdot & \cdot \end{pmatrix} \begin{pmatrix} A - \frac{1}{d}bc^T & b \\ \mathbf{0}^T & d \end{pmatrix} &= \begin{pmatrix} I_{N-1} & \mathbf{0} \\ -\frac{1}{d}c^T & 1 \end{pmatrix} \\ \Rightarrow \tilde{M} \begin{pmatrix} A - \frac{1}{d}bc^T \\ \mathbf{0}^T \end{pmatrix} &= I_{N-1} \\ \Rightarrow \tilde{M}^{-1} &= A - \frac{1}{d}bc^T. \end{aligned}$$

□

A.3 Calculating covariance matrices with an environmental cue

In this section, we show how to calculate covariance matrices when an environmental cue is present. In the text, we considered the noisy opinion dynamics given by

$$dv = Mv + DdW_t,$$

where M is a matrix such that $\mathbf{1}$ is an eigenvector with eigenvalue 0 and all other eigenvalues have negative real part and D is a diagonal matrix. Following Young et al. (2010) and Young (2014), we transformed the opinion vector v into a vector y that captures the component of v orthogonal to the consensus vector $\mathbf{1}$, using a matrix $Q \in \mathbb{R}^{(N-1) \times N}$ such that each row of Q was normal and orthogonal to $\mathbf{1}$. Specifically, we defined $\bar{v} = \sum_i v_i/N$, $y = Qv$, and $z = v - \bar{v}\mathbf{1} = Q^T y$. We then showed that y follows the dynamics

$$\begin{aligned} dy &= QMvdt + QDdW_t \\ &= QM(z + \bar{v}\mathbf{1})dt + QDdW_t \\ &= QMQ^T ydt + \bar{v}QM\mathbf{1}dt + QDdW_t \\ &= QMQ^T ydt + QDdW_t \\ &= \tilde{M}ydt + \tilde{D}dW_t. \end{aligned}$$

where $\tilde{M} = QMQ^T$ and $\tilde{D} = QD$. This allowed us to use the Lyapunov equation, (2), to find $\Sigma_{ss,y}$, as in Bamieh et al. (2012), Xiao et al. (2007), Young et al. (2010), Young (2014), and Zelazo and Mesbahi (2009).

Regardless of the choice of M and D , the stochastic opinion dynamics can incorporate an environmental cue, just as the deterministic opinion dynamics can. In this case,

$$dv = (M_s v + B\mathbf{1})dt + DdW_t$$

where $M_s = M - B$. A similar framework is used to study noisy opinion dynamics with an environmental cue in Fitch and Leonard (2014) and Pais and Leonard (2014). Now the dynamics of y follow

$$\begin{aligned} dy &= Q(M_s v + B\mathbf{1})dt + QDdW_t \\ &= Q(M_s(z + \bar{v}\mathbf{1}) + B\mathbf{1})dt + QDdW_t \\ &= (QM_s Q^T y - QB\bar{v}\mathbf{1} + QB\mathbf{1})dt + QDdW_t, \text{ since } M\mathbf{1} = \mathbf{0} \\ &= (QM_s Q^T y + (1 - \bar{v})QB\mathbf{1})dt + QDdW_t \\ &= (\tilde{M}_s y + (1 - \bar{v})\tilde{B}\mathbf{1})dt + \tilde{D}dW_t. \end{aligned}$$

In this case, the dynamics still tend to reduce deviations from consensus, but in a way that depends on \bar{v} . Without a cue, we could consider the $N - 1$ dimensional vector y on its own. With a cue, we need to consider a full N

dimensional vector that captures both y and \bar{v} . To do this, we use a full rotation matrix, $R \in \mathbb{R}^{N \times N}$ such that the first $N - 1$ rows of R are given by Q and the last is equal to $\frac{1}{\sqrt{N}}\mathbf{1}^T$. We now define $\bar{y} = Rv \in \mathbb{R}^N$. Note that $y = Qv$ is given by $y = (\bar{y}_1, \dots, \bar{y}_{N-1})^T$. By Ito's formula,

$$\begin{aligned} d\bar{y} &= (RM_s v + RB\mathbf{1})dt + RDdW_t \\ &= (RM_s R^T \bar{y} + RB\mathbf{1})dt + RDdW_t \\ &= (\tilde{M}_s \bar{y} + \tilde{B}\mathbf{1})dt + \tilde{D}dW_t \end{aligned}$$

where $\tilde{M}_s = RM_s R^T$, $\tilde{B} = RB$, and $\tilde{D} = RD$. If all the eigenvalues of M_s have negative real part, so too do all the eigenvalues of \tilde{M}_s , so that by Theorem A.2, $\Sigma_{ss, \bar{y}}$ satisfies the Lyapunov equation

$$\tilde{M}_s \Sigma_{ss, \bar{y}} + \Sigma_{ss, \bar{y}} \tilde{M}_s^T = -\tilde{D}\tilde{D}^T.$$

Pais and Leonard (2014) show that all the eigenvalues of M_s have negative real part if there is a path in the network from every node to at least one node that perceives the cue. In the text, we explain that if M has more than one zero eigenvalue, we set ρ equal to 0. We present no results about the covariance in the presence of the cue, but we could use a similar convention to define ρ in the case when M_s has a 0 eigenvalue. Once we solve this equation to find $\Sigma_{ss, \bar{y}}$, we can find $\Sigma_{ss, y}$ by taking the first $N - 1$ rows and columns. We can then find $\Sigma_{ss, z}$ using the transformation $\Sigma_{ss, z} = Q^T \Sigma_{ss, y} Q$. As before, we can use either $M = L_s$ or $M = P_s$ and either D with diagonal elements given by $\{\sqrt{k_1}, \dots, \sqrt{k_N}\}$ or $D = I_N$.

A.4 Calculating covariance matrices from the Boltzmann distribution

In this section, we describe how to find the covariance matrix Σ_z using the Boltzmann distribution. Since this framework requires a symmetric interaction network, we use the symmetrized matrix P .

First, we consider the Hamiltonian functions giving the energy of the opinion vector:

$$\begin{aligned} H(v) &= -v^T P v \text{ without the cue and} \\ H_s(v) &= -(v - \mathbf{1})^T P_s (v - \mathbf{1}) \text{ with the cue.} \end{aligned}$$

H is high when the opinions are far from consensus and low when the opinions are close to consensus. These functions let us write Boltzmann distributions over the opinion vector at steady state:

$$P(v) = \frac{1}{Z} \exp(-H(v)) \text{ without the cue and} \tag{7}$$

$$P_s(v) = \frac{1}{Z_s} \exp(-H_s(\bar{v})) \text{ with the cue,} \tag{8}$$

where $Z = \int_{\mathbb{R}^N} \exp(-H(v)) dv$ and $Z_s = \int_{\mathbb{R}^N} \exp(-H_s(v)) dv$.

We can then use these distributions to find the covariance matrix Σ_z . We will first use Eq. 7, which does not include an environmental cue. If $\bar{v} = \sum_i v_i / N$, then specifying z and \bar{v} is equivalent to specifying v , which lets us find the probability distribution over z :

$$\begin{aligned} P(v) &= \frac{1}{Z} \exp\left(-\frac{1}{2}(z + \bar{v}\mathbf{1})^T (-P)(z + \bar{v}\mathbf{1})\right) \\ &= \frac{1}{Z} \exp\left(-\frac{1}{2}z^T (-P)z\right) \text{ since } P\mathbf{1} = P^T \mathbf{1} = \mathbf{0} \\ \Rightarrow P(z) &\propto \exp\left(\frac{1}{2}z^T (-P)z\right). \end{aligned}$$

Since P has a 0 eigenvalue, it is not invertible. However, z is orthogonal to $\mathbf{1}$, the eigenvector of P with eigenvalue 0. We can therefore discard the 0 eigenvalue (Bialek et al. 2012). P is diagonalizable since it is symmetric, so if Λ is the diagonal matrix with eigenvalues of P and W is the matrix whose columns are the eigenvectors of P , then $P = W\Lambda W^T$. If $\lambda_1 = 0$, then let $\tilde{\Lambda}$ be the $(N - 1) \times (N - 1)$ diagonal matrix with $\lambda_2, \dots, \lambda_N$ along the diagonal and \tilde{W} be W with the first column removed. We can now invert the non-zero eigenvalues of P and define $P^\dagger = \tilde{W}\tilde{\Lambda}^{-1}\tilde{W}^T$. Finally, the covariance of z is $\Sigma_z = -P^\dagger$ and of y is $\Sigma_y = -QP^\dagger Q^T$. By dividing each entry Σ_{zij} of the covariance matrix by $\sqrt{\Sigma_{zii}\Sigma_{zjj}}$, the covariance matrix can be used to find the correlation matrix C_z .

In the above, we assumed that P has only one 0 eigenvalue. The topology of the network captured by S is undirected, since wherever there is an edge going one way there is an edge going the other. In an undirected network, the multiplicity of the 0 eigenvalue of the Laplacian is equal to the number of connected components (Olfati-Saber and Murray 2004). Suppose there are m connected components, $\mathcal{C}^1, \dots, \mathcal{C}^m$. Then the vector w^j defined by

$$w_i^j = \begin{cases} 1 & \text{if } i \in \mathcal{C}^j \\ 0 & \text{otherwise} \end{cases}$$

is an eigenvector of P with eigenvalue 0 (Olfati-Saber and Murray 2004; Pais and Leonard 2014). If $i \in \mathcal{C}^j$ and

$$\bar{v}^j = \frac{\sum_{k \in \mathcal{C}^j} v_k}{\# \text{ of nodes in } \mathcal{C}^j},$$

instead of defining $z_i = v_i - \bar{v}$ we can define $z_i = v_i - \bar{v}^j$. Then z is orthogonal to each of the m vectors, w^j . This allows us to discard each of the m eigenvectors with 0 eigenvalue, just as we discarded the single 0 eigenvector above.

We will now use Eq. 8 to find Σ_z when an environmental cue is present. As before, let $\bar{v} = \sum_i v_i/N$ and $z = v - \bar{v}\mathbf{1}$.

$$P_s(v) = \frac{1}{Z} \exp\left(-\frac{1}{2}(v - \mathbf{1})^T(-P_s)(v - \mathbf{1})\right) = \frac{1}{Z} \exp\left(-\frac{1}{2}(z - (1 - \bar{v})\mathbf{1})^T(-P_s)(z - (1 - \bar{v})\mathbf{1})\right)$$

Since the subspace of $z \in \mathbb{R}^N$ such that $\sum_i z_i = 0$ is $N - 1$ dimensional, it will be easier to rotate z into \mathbb{R}^{N-1} . Therefore, we consider $\bar{y} = Rz$. Since $\sum_i z_i = 0$, $z \cdot \mathbf{1} = 0$ so that $\bar{y}_N = 0$. Note that if $e_N = (0, \dots, 0, 1)^T$, then $R\mathbf{1} = \sqrt{N}e_N$ so that $\mathbf{1} = \sqrt{N}R^T e_N$. This allows us to write

$$P(v) = \frac{1}{Z} \exp\left(-\frac{1}{2}(R^T \bar{y} - (1 - \bar{v})\sqrt{N}R^T e_N)^T(-P_s) \times (R^T \bar{y} - (1 - \bar{v})\sqrt{N}R^T e_N)\right) = \frac{1}{Z} \exp\left(-\frac{1}{2}(\bar{y} - (1 - \bar{v})\sqrt{N}e_N)^T(-RP_s R^T) \times (\bar{y} - (1 - \bar{v})\sqrt{N}e_N)\right) = \frac{1}{Z} \exp\left(-\frac{1}{2}(\bar{y} + \sqrt{N}\bar{v}e_N - \sqrt{N}e_N)^T(-\tilde{P}_s) \times (\bar{y} + \sqrt{N}\bar{v}e_N - \sqrt{N}e_N)\right)$$

where $\tilde{P}_s = RP_s R^T$. To find $P(\bar{y})$ we can integrate out \bar{v} . However, to consider this as a probability density function over \bar{y} and \bar{v} , we have to consider the determinant of the transformation from v to \bar{y} and \bar{v} . Since $\bar{v} = 1/N\mathbf{1}^T v = 1/\sqrt{N}R_N^T v$ where R_N is the last row of R , the transformation can be represented by $\bar{R} = \begin{pmatrix} Q \\ 1/\sqrt{N}R_N \end{pmatrix}$. The determinant of this transformation is given by $\det(\bar{R}) = \frac{1}{\sqrt{N}} \det(R) = \frac{1}{\sqrt{N}}$. Therefore, $P(\bar{y}, \bar{v}) = \sqrt{N}P(v)$ is the

appropriate density function for \bar{y} and \bar{v} . Finally, note that $Z = \sqrt{(2\pi)^N \det(-P_s^{-1})}$. This gives

$$P(\bar{y}, \bar{v}) = \sqrt{N}P(v) = \frac{\sqrt{N}}{Z} \exp\left(-\frac{1}{2}((\bar{v} - 1)\sqrt{N}e_N + \bar{y})^T(-\tilde{P}_s) \times ((\bar{v} - 1)\sqrt{N}e_N + \bar{y})\right) \Rightarrow P(\bar{y}) = \int_{\mathbb{R}} P(\bar{y}, \bar{v})d\bar{v} = \frac{\sqrt{N}}{Z} \exp\left(-\frac{1}{2}\bar{y}^T(-\tilde{P}_s)\bar{y}\right) \times \int_{\mathbb{R}} \exp\left(-\frac{1}{2}\left(-\tilde{P}_{sNN}N(\bar{v} - 1)^2 - (\bar{v} - 1)\sqrt{N}\tilde{P}_{sN} \cdot \bar{y} - (\bar{v} - 1)\sqrt{N}\bar{y}^T \tilde{P}_{sN}\right)\right)d\bar{v} = \frac{\sqrt{N}}{Z} \exp\left(-\frac{1}{2}\bar{y}^T(-\tilde{P}_s)\bar{y}\right) \times \int_{\mathbb{R}} \exp\left(-\frac{1}{2}\left(-\tilde{P}_{sNN}N(\bar{v} - 1)^2 - 2(\bar{v} - 1)\sqrt{N}\tilde{P}_{sN} \cdot \bar{y}\right)\right)d\bar{v} \text{ since } \tilde{P}_s \text{ is symmetric} = \frac{\sqrt{N}}{Z} \exp\left(-\frac{1}{2}\left(\bar{y}^T(-\tilde{P}_s)\bar{y} - \left(\frac{\tilde{P}_{sN} \cdot \bar{y}}{\sqrt{-\tilde{P}_{sNN}}}\right)^2\right)\right) \times \int_{\mathbb{R}} \exp\left(-\frac{1}{2}\left(-\tilde{P}_{sNN}N(\bar{v} - 1)^2 - 2(\bar{v} - 1)\sqrt{N}\tilde{P}_{sN} \cdot \bar{y} + \left(\frac{\tilde{P}_{sN} \cdot \bar{y}}{\sqrt{-\tilde{P}_{sNN}}}\right)^2\right)\right)d\bar{v} = \frac{\sqrt{N}}{Z} \exp\left(-\frac{1}{2}\left(\bar{y}^T\left(-\tilde{P}_s + \frac{1}{\tilde{P}_{sNN}}\tilde{P}_{sN}\tilde{P}_{sN}\right)\bar{y}\right)\right) \times \int_{\mathbb{R}} \exp\left(-\frac{1}{2}\left(\sqrt{-\tilde{P}_{sNN}N}(\bar{v} - 1) - \frac{\tilde{P}_{sN} \cdot \bar{y}}{\sqrt{-\tilde{P}_{sNN}}}\right)^2\right)d\bar{v} = \frac{\sqrt{N}}{Z} \exp\left(-\frac{1}{2}\left(\bar{y}^T\left(-\tilde{P}_s + \frac{1}{\tilde{P}_{sNN}}\tilde{P}_{sN}\tilde{P}_{sN}\right)\bar{y}\right)\right) \times \int_{\mathbb{R}} \exp\left(-\frac{1}{2}\left(-\tilde{P}_{sNN}N\left(\bar{v} - 1 + \frac{\tilde{P}_{sN} \cdot \bar{y}}{\tilde{P}_{sNN}\sqrt{N}}\right)\right)^2\right)d\bar{v} = \frac{\sqrt{N}}{Z} \exp\left(-\frac{1}{2}\left(\bar{y}^T\left(-\tilde{P}_s + \frac{1}{\tilde{P}_{sNN}}\tilde{P}_{sN}\tilde{P}_{sN}\right)\bar{y}\right)\right) \times \sqrt{\frac{2\pi}{-\tilde{P}_{sNN}N}} = \frac{1}{\sqrt{(2\pi)^{N-1} \det(-P_s^{-1})(-\tilde{P}_{sNN})}} \exp\left(-\frac{1}{2}\left(\bar{y}^T\left(-\tilde{P}_s + \frac{1}{\tilde{P}_{sNN}}\tilde{P}_{sN}\tilde{P}_{sN}\right)\bar{y}\right)\right)$$

Therefore, the covariance of \bar{y} is $\Sigma_{\bar{y}} = (-\tilde{P}_s + \frac{1}{\tilde{P}_{sNN}}\tilde{P}_{sN}\tilde{P}_{sN})^{-1}$. Then, to rotate back into z we use $z = R^T \bar{y}$ so that

$$\begin{aligned} \Sigma_z &= R^T \left(-\tilde{P}_s + \frac{1}{\tilde{P}_{sNN}}\tilde{P}_{sN}\tilde{P}_{sN} \right)^{-1} R \\ \Rightarrow \Sigma_z^{-1} &= R^T \left(-\tilde{P}_s + \frac{1}{\tilde{P}_{sNN}}\tilde{P}_{sN}\tilde{P}_{sN} \right) R \\ &= -R^T R P_s R^T R + \frac{1}{\tilde{P}_{sNN}} R^T R P_s R_{sN}^T R_{sN} P_s R^T R \\ &= -P_s + \frac{1}{N\tilde{P}_{sNN}} P_s \mathbf{1}\mathbf{1}^T P_s \\ &= -P_s + \frac{1}{N\tilde{P}_{sNN}} B \mathbf{1}\mathbf{1}^T B \\ \Rightarrow \Sigma_z &= \left(-P_s + \frac{1}{N\tilde{P}_{sNN}} B \mathbf{1}\mathbf{1}^T B \right)^{-1} \end{aligned}$$

Let N_c be the number of birds that perceive the cue. Then,

$$\tilde{P}_{sNN} = -\frac{1}{N} \mathbf{1}^T B \mathbf{1} = -\frac{bN_c}{N}.$$

Finally, we find that

$$\Sigma_z = -\left(P_s + \frac{1}{bN_c} B \mathbf{1}\mathbf{1}^T B \right)^{-1}$$

Note that

$$\begin{aligned} \left(P_s + \frac{1}{bN_c} B \mathbf{1}\mathbf{1}^T B \right) \mathbf{1} &= -B \mathbf{1} + \frac{1}{bN_c} B \mathbf{1}\mathbf{1}^T B \mathbf{1} \\ &= -B \mathbf{1} + \frac{bN_c}{bN_c} B \mathbf{1} = \mathbf{0}, \end{aligned}$$

so we cannot actually invert the matrix we want to. However, as above, we can discard the 0 eigenvector(s). This gives us $\Sigma_z = -(P_s + \frac{1}{bN_c} B \mathbf{1}\mathbf{1}^T B)^\dagger$ and $\Sigma_y = -Q(P_s + \frac{1}{bN_c} B \mathbf{1}\mathbf{1}^T B)^\dagger Q^T$. Again, we divide each Σ_{zij} by $\sqrt{\Sigma_{zii} \Sigma_{zjj}}$ to find the correlation matrix C_z .

To summarize, by using the Boltzmann distribution, we find that in the absence of an environmental cue $\Sigma_z = -P^\dagger$, and in the presence of a cue, $\Sigma_z = -(P_s + \frac{1}{bN_c} B \mathbf{1}\mathbf{1}^T B)^\dagger$. In each case, the correlation matrix is found by $C_{zij} = \Sigma_{zij} / \sqrt{\Sigma_{zii} \Sigma_{zjj}}$.

A.5 Equivalence of two steady state distributions

In this section, we compare the two approaches, the Lyapunov equations arising from a set of stochastic differential equations, described in the text and in Appendix A.3, and the Boltzmann distribution, described in Appendix A.4.

The two approaches can be used with and without an environmental cue and the resulting covariance matrices Σ_y are summarized in Table 2. The interaction networks described in L , arising from topological neighborhoods, are typically not symmetric. Since the Boltzmann distribution requires a symmetric interaction matrix, the Lyapunov equation method is a bit more general. Below, we show that if the same symmetric interaction matrix is used in both approaches, we find the same covariance matrices up to a factor of 2. According to Theorem A.2, if $dv = Mvdt + DdW_t$, then the probability densities of v , $y = Qv$, and $z = Q^T y$ are given by Gaussian distributions. Since the Boltzmann distribution is a Gaussian distribution, it makes sense that the steady state distribution from the stochastic dynamics and the Boltzmann distribution should be the same. In Fig. 7, we show that even if we use different interaction matrices, the two approaches give very similar covariance matrices. The slight differences in Fig. 7 are due to the fact that L and P are slightly different, since L is not symmetric.

Since the Boltzmann distribution requires a symmetric interaction matrix, in the following we will only consider the interaction matrix P . We start by considering the case without an environmental cue. As we showed in the main text, if $dv = Pvd t + dW_t$, then $\Sigma_{ss,y}$ satisfies the equation

$$\bar{P} \Sigma_{ss,y} + \Sigma_{ss,y} \bar{P} = -Q Q^T,$$

where $\bar{P} = Q P Q^T$. If, on the other hand, we start with the Boltzmann distribution, we find that $\Sigma_z = -P^\dagger$ and, since $y = Qz$, $\Sigma_y = -Q P^\dagger Q^T$.

Claim 1 If $\Sigma_y^* = -Q P^\dagger Q^T$ and Σ_y satisfies the equation $\bar{P} \Sigma_y + \Sigma_y \bar{P} = -Q Q^T$, then $\Sigma_y = \frac{1}{2} \Sigma_y^*$.

Proof We begin by using Σ_y^* in the Lyapunov equation:

$$\begin{aligned} \bar{P} \Sigma_y^* + \Sigma_y^* \bar{P} &= Q P Q^T (-Q P^\dagger Q^T) + (-Q P^\dagger Q^T) Q P^T \\ &= -Q P \left(I_N - \frac{1}{N} \mathbf{1}\mathbf{1}^T \right) P^\dagger Q^T \\ &\quad - Q P^\dagger \left(I_N - \frac{1}{N} \mathbf{1}\mathbf{1}^T \right) P^T \\ &= -Q P P^\dagger Q^T - Q P^\dagger P^T Q^T \text{ since } P \mathbf{1} = \mathbf{0} \\ &= -Q \left(I_N - \frac{1}{N} \mathbf{1}\mathbf{1}^T \right) Q^T \\ &\quad - Q \left(I_N - \frac{1}{N} \mathbf{1}\mathbf{1}^T \right) Q^T \end{aligned}$$

$$\begin{aligned} &= -QQ^T - QQ^T \text{ since } Q\mathbf{1} = \mathbf{0} \\ &= -2QQ^T \end{aligned}$$

Thus, $\frac{1}{2}\Sigma_y^*$ satisfies the Lyapunov equation. Since Theorem A.1 tells us that the solution is unique, $\Sigma_y = \frac{1}{2}\Sigma_y^*$. \square

We can now consider the case when there is an environmental cue. In Appendix A.3, we used Theorem A.2 to conclude that, if $dv = (P_s v + B\mathbf{1})dt + dW_t$, then the covariance of the full vector $\bar{y} = Rv, \Sigma_{ss,\bar{y}}$, satisfies the equation

$$\tilde{P}\Sigma_{ss,\bar{y}} + \Sigma_{ss,\bar{y}}\tilde{P} = -I,$$

where $\tilde{P} = RPR^T$. If, on the other hand, we start with the Boltzmann distribution $P(v) \propto \exp(-\frac{1}{2}(v - \mathbf{1})(-P_s)(v - \mathbf{1}))$, we find that $\Sigma_z = -(P_s + \frac{1}{bN_c}B\mathbf{1}\mathbf{1}^T B)^\dagger$ where N_c is the number of birds that perceive the cue.

Claim A.2 Let $\Sigma_z^* = -(P_s + \frac{1}{bN_c}B\mathbf{1}\mathbf{1}^T B)^\dagger$ and $\Sigma_y^* = Q\Sigma_z^*Q^T$. If $\Sigma_{\bar{y}}$ satisfies the Lyapunov equation $\tilde{P}_s\Sigma_{\bar{y}} + \Sigma_{\bar{y}}\tilde{P}_s = -I$ and Σ_y is given by the first $N - 1$ rows and columns of $\Sigma_{\bar{y}}$, then $\Sigma_y = \frac{1}{2}\Sigma_y^*$.

Proof It is clear that $-\frac{1}{2}\tilde{P}_s^{-1}$ satisfies the Lyapunov equation. By Theorem A.1 the solution to the Lyapunov equation is unique, so $\Sigma_{\bar{y}} = -\frac{1}{2}\tilde{P}_s^{-1}$. Equivalently, we can write

$$\Sigma_{\bar{y}}^{-1} = -2\tilde{P}_s = -2RP_sR^T.$$

In order to use Claim A.2, we will write $\Sigma_{\bar{y}}^{-1}$ as

$$\begin{aligned} -2RP_sR^T &= -2 \begin{pmatrix} QP_sQ^T & \frac{1}{\sqrt{N}}QP_s\mathbf{1} \\ \frac{1}{\sqrt{N}}\mathbf{1}^T P_sQ^T & \frac{1}{N}\mathbf{1}^T P_s\mathbf{1} \end{pmatrix} \\ &\quad \text{since the last row of } R \text{ is } \frac{1}{\sqrt{N}}\mathbf{1}^T \\ &= -2 \begin{pmatrix} QP_sQ^T & -\frac{1}{\sqrt{N}}QB\mathbf{1} \\ -\frac{1}{\sqrt{N}}\mathbf{1}^T BQ^T & -\frac{1}{N}\mathbf{1}^T B\mathbf{1} \end{pmatrix} \\ &\quad \text{since } P_s = P - B \text{ and } P\mathbf{1} = \mathbf{0} \\ &= -2 \begin{pmatrix} QP_sQ^T & -\frac{1}{\sqrt{N}}QB\mathbf{1} \\ -\frac{1}{\sqrt{N}}\mathbf{1}^T BQ^T & -\frac{1}{bN_c} \end{pmatrix} \end{aligned}$$

Now, by Claim A.2, since Σ_y is given by the first $N - 1$ rows and columns of $\Sigma_{\bar{y}}$,

$$\begin{aligned} \Sigma_y^{-1} &= -2 \left(QP_sQ^T + \frac{N}{bN_c} \frac{1}{N} QB\mathbf{1}\mathbf{1}^T BQ^T \right) \\ &= -2Q \left(P_s + \frac{1}{bN_c} B\mathbf{1}\mathbf{1}^T B \right) Q^T. \end{aligned}$$

Since Q and Q^T are not square, they are not invertible; nor is $P_s + \frac{1}{bN_c}B\mathbf{1}\mathbf{1}^T B$, so it takes some care to find Σ_y from Σ_y^{-1} . Let $M = -\frac{1}{2}Q \left(P_s + \frac{1}{bN_c}B\mathbf{1}\mathbf{1}^T B \right)^\dagger Q^T$. Then

$$\begin{aligned} \Sigma_y^{-1}M &= \left(-2Q \left(P_s + \frac{1}{bN_c} B\mathbf{1}\mathbf{1}^T B \right) Q^T \right) \\ &\quad \times \left(-\frac{1}{2}Q \left(P_s + \frac{1}{bN_c} B\mathbf{1}\mathbf{1}^T B \right)^\dagger Q^T \right) \\ &= Q \left(P_s + \frac{1}{bN_c} B\mathbf{1}\mathbf{1}^T B \right) Q^T Q \left(P_s + \frac{1}{bN_c} B\mathbf{1}\mathbf{1}^T B \right)^\dagger Q^T \\ &= Q \left(P_s + \frac{1}{bN_c} B\mathbf{1}\mathbf{1}^T B \right) \left(I_N - \frac{1}{N} \mathbf{1}\mathbf{1}^T \right) \\ &\quad \times \left(P_s + \frac{1}{bN_c} B\mathbf{1}\mathbf{1}^T B \right)^\dagger Q^T \\ &= Q \left(P_s \left(I_N - \frac{1}{N} \mathbf{1}\mathbf{1}^T \right) + \frac{1}{bN_c} B\mathbf{1}\mathbf{1}^T B \right) \\ &\quad \times \left(I_N - \frac{1}{N} \mathbf{1}\mathbf{1}^T \right) \left(P_s + \frac{1}{bN_c} B\mathbf{1}\mathbf{1}^T B \right)^\dagger Q^T \\ &= Q \left(P_s - \frac{1}{N} B\mathbf{1}\mathbf{1}^T + \frac{1}{bN_c} B\mathbf{1}\mathbf{1}^T B \right) \\ &\quad - \frac{1}{N} \frac{1}{bN_c} B\mathbf{1}\mathbf{1}^T B\mathbf{1}\mathbf{1}^T \left(P_s + \frac{1}{bN_c} B\mathbf{1}\mathbf{1}^T B \right)^\dagger Q^T \\ &\quad \text{since } P\mathbf{1} = \mathbf{0} \\ &= Q \left(P_s - \frac{1}{N} B\mathbf{1}\mathbf{1}^T + \frac{1}{bN_c} B\mathbf{1}\mathbf{1}^T B \right) \\ &\quad - \frac{1}{N} \frac{bN_c}{bN_c} B\mathbf{1}\mathbf{1}^T \left(P_s + \frac{1}{bN_c} B\mathbf{1}\mathbf{1}^T B \right)^\dagger Q^T \\ &= Q \left(P_s + \frac{1}{bN_c} B\mathbf{1}\mathbf{1}^T B \right) \left(P_s + \frac{1}{bN_c} B\mathbf{1}\mathbf{1}^T B \right)^\dagger Q^T \\ &= Q \left(I_N - \frac{1}{N} \mathbf{1}\mathbf{1}^T \right) Q^T \\ &= QQ^T \text{ since } Q\mathbf{1} = \mathbf{0} \\ &= I_{N-1}. \end{aligned}$$

Finally, we can conclude that $\Sigma_y = -\frac{1}{2}Q \left(P_s + \frac{1}{bN_c} B\mathbf{1}\mathbf{1}^T B \right)^\dagger Q^T = \frac{1}{2}Q\Sigma_z^*Q^T = \frac{1}{2}\Sigma_y^*$. \square

A.6 Supplemental figures and tables

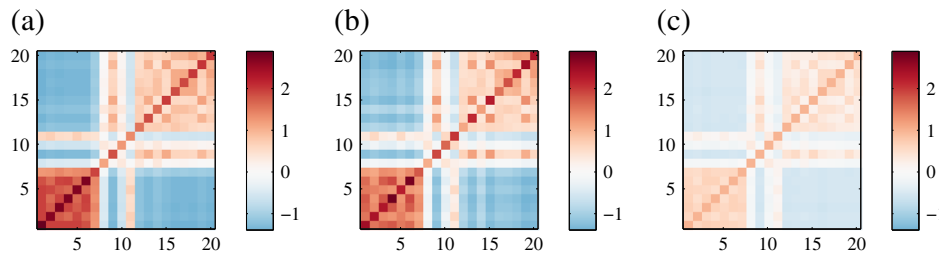


Fig. 7 Example covariance matrices. We determine L and P by distributing the birds randomly in the unit square and having each pay attention to its 4 nearest neighbors. In **a** we show the covariance matrix $\frac{1}{2}Q^T \Sigma_y Q$ where Σ_y is the solution to the Lyapunov equation (2) using

L . In **b** we show the covariance matrix $\Sigma_z = -P^\dagger$. This shows that $\Sigma_z \approx \frac{1}{2}Q^T \Sigma_y Q$. In **c** we show the correlation matrix C found by normalizing the entries of Σ_z

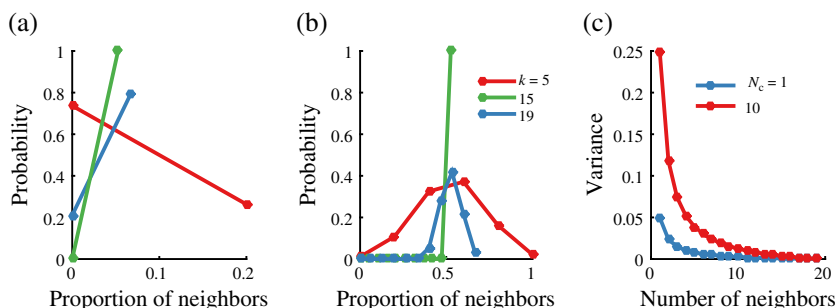


Fig. 8 If a fixed number of birds perceive the cue and birds choose neighbors randomly, the probability of a proportion of neighbors that perceive the cue follows Eq. 3. In **a** and **b** we show the probability of having a given proportion of neighbors for different strategies k . In **a**,

$N_c = 1$. In **b**, $N_c = 10$. In **c**, we show the variance of the proportion of neighbors that perceive the cue, as given in Eq. 5, as a function of the number of neighbors, for $N_c = 1$ and $N_c = 10$. Parameters: $N = 20$

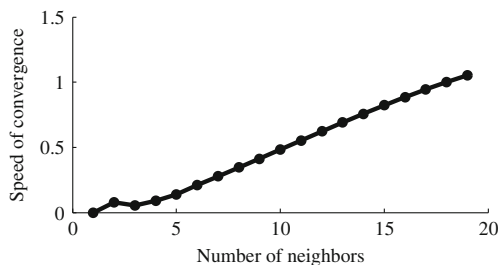


Fig. 9 Speed of convergence increases with number of neighbors. Here we show the average speed of convergence σ over 1000 cues, where all birds use the strategy indicated on the x-axis. Parameters: $N = 20$

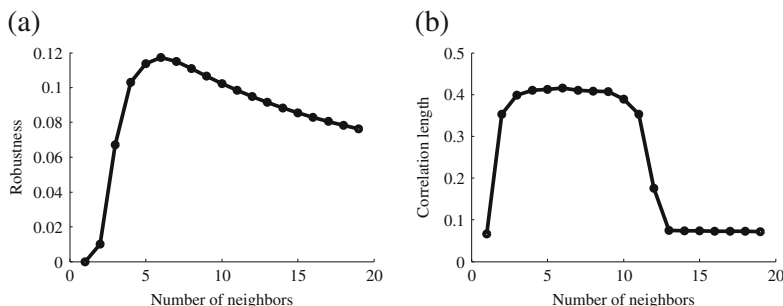


Fig. 10 Both robustness and correlation length are maximized at intermediate numbers of neighbors. In **a**, we show the average robustness, $\bar{\rho}$, from 1000 cues, where all birds use the strategy indicated on the

x-axis. In **b**, we show the correlation length, λ , found using the correlation matrices from 1000 cues, where all birds use the strategy indicated on the x-axis. Parameters: $N = 20$

Table 2 The table shows how to calculate the steady state covariance matrix Σ_y

	Without cue	With cue
Stochastic dynamics	$\tilde{P}\Sigma_y + \Sigma_y\tilde{P} = -Q Q^T$	$\tilde{P}_s\Sigma_{\tilde{y}} + \Sigma_{\tilde{y}}\tilde{P}_s = -I_N, \Sigma_y = (\Sigma_{\tilde{y}})_{i,j=1}^{N-1}$
Boltzmann	$\Sigma_y = -Q P^\dagger Q^T$	$\Sigma_y = -Q(P_s + \frac{1}{bN_c} B \mathbf{1}\mathbf{1}^T B)^\dagger Q^T$

The probability density of the opinion vector can come from analyzing the stochastic dynamics, $dv = Pdt + dW_t$ without a cue and $dv = (P_s + B\mathbf{1})dt + dW_t$ with a cue, or from analyzing the Boltzmann distributions, $P(v) \propto \exp(-\frac{1}{2}v^T(-P)v)$ without a cue and $P(v) \propto \exp(-\frac{1}{2}(v - \mathbf{1})^T(-P_s)(v - \mathbf{1}))$ with a cue. For any matrix M , $\tilde{M} = QMQ^T$, $\tilde{M} = RMR^T$, and M^\dagger is the inverse of M with the 0 eigenvector and eigenvalue removed

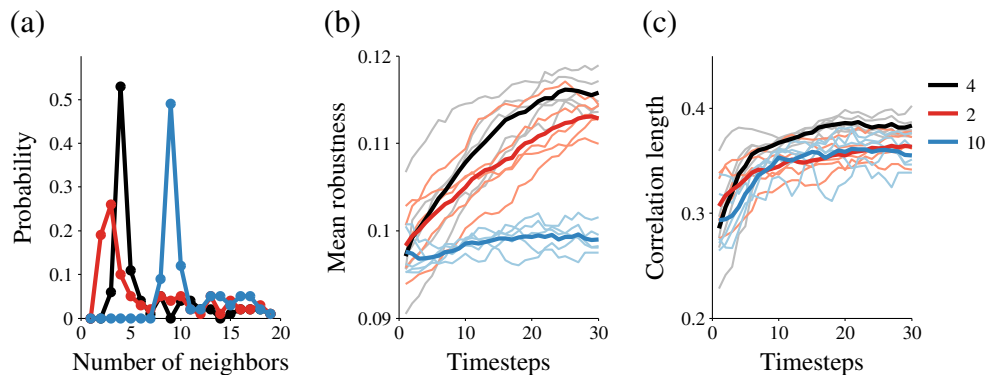


Fig. 11 The greedy optimization process is similar for different numbers of birds that are allowed to eat. In the main text, we showed results for fitness based on access to abundant resources, when four birds in the flock are allowed to eat. Here we show results for when two, four, or ten birds in the flock are allowed to eat. For each of these cases, we ran a greedy optimization process with five random initial conditions. In each subplot, *red lines* indicate that two birds are allowed to eat, *black lines* indicate that four birds are allowed to eat, and *blue lines* indicate that ten birds are allowed to eat. In **a**, we show the probability of a bird ending with each strategy, $k = 1, \dots, N - 1$, after 30 timesteps, calculated by the number of times a bird ended with each strategy, divided by $5N$. As more birds are allowed to eat, higher

strategies become more probable. In **b**, we show, for various initial conditions, how the average robustness, $\bar{\rho}$, changes over time, with the *darker lines* showing the average of the 5 paler lines of the same color. The average robustness tends to be highest when four birds are allowed to eat, but in neither of the other two cases does the average robustness decrease as much as when one bird is allowed to eat (i.e. fitness is based on access to scarce resources, as defined in the main text). In **c**, we show, for various initial conditions, how the correlation length, λ , changes over time, with the *darker lines* showing the average of the 5 paler lines of the same color. The correlation length tends to be highest when four birds are allowed to eat, but it is almost as high in the other two cases. Parameters: $N = 20, b = 1, r = 0.1$

References

Agavev R, Chebotarev P (2005) On the spectra of nonsymmetric laplacian matrices. *Linear Algebra Appl* 399:157–168
 Arnold L (1974) *Stochastic differential equations: theory and applications*. Wiley
 Attanasi A, Cavagna A, Castello LD, Giardina I, Grigera TS, Jelic A, Melillo S, Parisi L, Pohl O, Shen E, Viale M (2014) Information transfer and behavioural inertia in starling flocks. *Nat Phys* 10(9):691–696
 Bak P, Tang C, Wiesenfeld K (1988) Self-organized criticality. *Phys Rev A* 38(1):364
 Ballerini M, Cabibbo N, Candelier R, Cavagna A, Cisbani E, Giardina I, Lecomte V, Orlandi A, Parisi G, Procaccini A, Viale M, Zdravkovic V (2008) Interaction ruling animal collective behavior depends on topological rather than metric distance: evidence from a field study. *Proc Natl Acad Sci* 105(4):1232–1237
 Bamieh B, Jovanovic MR, Mitra P, Patterson S (2012) Coherence in large-scale networks: dimension-dependent limitations of local feedback. *IEEE Trans Autom Control* 57(9):2235–2249
 Berger J (1978) Group size, foraging, and antipredator ploys: an analysis of bighorn sheep decisions. *Behav Ecol Sociobiol* 4(1):91–99

Bialek W, Cavagna A, Giardina I, Mora T, Silvestri E, Viale M, Walczak AM (2012) Statistical mechanics for natural flocks of birds. *Proc Natl Acad Sci* 109(13):4786–4791
 Bialek W, Cavagna A, Giardina I, Mora T, Pohl O, Silvestri E, Viale M, Walczak AM (2014) Social interactions dominate speed control in driving natural flocks toward criticality. *Proc Natl Acad Sci* 111(20):7212–7217
 Bode NW, Franks DW, Wood AJ (2010) Limited interactions in flocks: relating model simulations to empirical data. *J R Soc Interface*:rsfi20100397
 Cavagna A, Cimarelli A, Giardina I, Parisi G, Santagati R, Stefanini F, Viale M (2009) Scale-free correlations in starling flocks. *Proc Natl Acad Sci* 107(26):11,685–11,870
 Dall SRX, Giraldeau LA, Olsson O, McNamara JM, Stephens DW (2005) Information and its use by animals in evolutionary ecology. *TRENDS in Ecology and Evolution* 20(4):187–193
 Danchin É, Giraldeau LA, Valone TJ, Wagner RH (2004) Public information: from nosy neighbors to cultural evolution. *Science* 305(5683):487–491
 Dechmann DKN, Heucke SL, Giuggioli L, Safi K, Voigt CC, Wikelski M (2009) Experimental evidence for group hunting via eavesdropping in echolocating bats. *Proc R Soc Lond B Biol Sci* 276:2721–2728

- Dullerud GE, Paganini F (2000) A course in robust control theory. Springer
- Fitch KE, Leonard NE (2014) Joint centrality distinguishes optimal leaders in noisy networks. arXiv:1407(1569)
- Giraldeau LA, Dubois F (2008) Social foraging and the study of exploitative behavior. *Adv Study Behav* 38:59–104
- Giraldeau LA, Valone TJ, Templeton JJ (2002) Potential disadvantages of using socially acquired information. *Philos Trans R Soc Lond B* 357(1427):1559–1566
- Henrich J, Boyd R (1998) The evolution of conformist transmission and the emergence of between-group differences. *Evol Hum Behav* 19(4):215–241
- Katz Y, Tunström K, Ioannou CC, Huepe C, Couzin ID (2011) Inferring the structure and dynamics of interactions in schooling fish. *Proc Natl Acad Sci* 108(46):18,720–18,725
- King AJ, Cowlshaw G (2007) When to use social information: the advantage of large group size in individual decision making. *Biol Lett* 3(2):137–139
- Landeau L, Terborgh J (1986) Oddity and the 'confusion effect' in predation. *Anim Behav* 34(5):1372–1380
- Lukeman R, Li YZ, Edelstein-Keshet L (2010) Inferring individual rules from collective behavior. *Proc Natl Acad Sci* 107(28):12,576–12,580
- Molvar EM, Bowyer RT (1994) Costs and benefits of group living in a recently social ungulate: the Alaskan moose. *Journal of Mammology* 75(3):621–630
- Nowak M (2006) Evolutionary dynamics. Belknap / Harvard
- Olfati-Saber R, Murray RM (2004) Consensus problems in networks of agents with switching topology and time delays. *IEEE Trans Autom Control* 49(9):1520–1533
- Pais D, Leonard NE (2014) Adaptive network dynamics and evolution of leadership in collective migration. *Physica D: Nonlinear Phenomena* 267:81–93
- Pitcher T, Magurran A, Winfield I (1982) Fish in larger shoals find food faster. *Behav Ecol Sociobiol* 10(2):149–151
- Porter JM, Sealy SG (1982) Dynamics of seabird multispecies feeding flocks: age-related feeding behaviour. *Behaviour* 81(2):91–109
- Powell GVN (1974) Experimental analysis of the social value of flocking by starlings (*Strunus vulgaris*) in relation to predation and foraging. *Anim Behav* 22(2):501–505
- Rieucou G, Giraldeau LA (2011) Exploring the costs and benefits of social information use: an appraisal of current experimental evidence. *Philos Trans R Soc Lond B* 366(1567):949–957
- Rosenthal SB, Twomey CR, Hartnett AT, Wu HS, Couzin ID (2015) Revealing the hidden networks of interaction in mobile animal groups allows prediction of complex behavioral contagion. *Proc Natl Acad Sci* 112(15):4690–4695
- Shaw AK, Couzin ID (2013) Migration or residency? The evolution of movement behavior and information usage in seasonal environments. *Am Nat* 181(1):114–124
- Strandburg-Peshkin A, Twomey CR, Bode NW, Kao AB, Katz Y, Ioannou CC, Rosenthal SB, Torney CJ, Wu HS, Levin SA, Couzin ID (2013) Visual sensory networks and effective information transfer in animal groups. *Curr Biol* 23(17):R709–R711
- Strandburg-Peshkin A, Farine DR, Couzin ID, Crofoot MC (2015) Shared decision-making drives collective movement in wild baboons. *Science* 248(6241):1358–1361
- Sullivan KA (1984) The advantages of social foraging in downy woodpeckers. *Anim Behav* 32(1):16–22
- Torney CJ, Lorenzi T, Couzin ID, Levin SA (2015) Social information use and the evolution of unresponsiveness in collective systems. *J R Soc Interface* 12(103):20140,893
- Vickery WL, Giraldeau LA, Templeton JJ, Kramer DL, Chapman CA (1991) Producers, scroungers, and group foraging. *Am Nat*:847–863
- Weimerskirch H, Bertrand S, Silva J, Marques JC, Goya E (2010) Use of social information in seabirds: compass rafts indicate the heading of food patches. *PLoS One*:e9928
- Xiao L, Boyd S, Kim SJ (2007) Distributed average consensus with least-mean square deviation. *J Parallel Distrib Comput* 67(1):33–46
- Young GF (2014) Optimising robustness of consensus to noise on directed networks. PhD thesis, Princeton University
- Young GF, Scardovi L, Leonard NE (2010) Robustness of noisy consensus dynamics with directed communication. In: Proceedings of the american control conference, pp 6312–6317
- Young GF, Scardovi L, Cavagna A, Giardina I, Leonard NE (2013) Starling flock networks manage uncertainty in consensus at low cost. *PLoS Computational Biology* 9(1):e1002,894
- Zelazo D, Mesbahi M (2009) H2 performance of agreement protocol with noise: an edge based approach. In: Proceedings of the IEEE conference on decision and control, pp 4747–4752