Evolution of Information Gathering Strategies

Contents

1 Introduction

Questions:

1. How are the two group properties—$\mathscr{H}_2$ robustness and correlation length—related?

2. What are the equilibrium strategies of individuals seeking to optimize their own fitness?

3. What group properties emerge from individual-level optimization?

2 Model

2.1 Modeling interactions and group-level properties There are two ways to describe equilibrium properties of a flock sharing information over a network. The first is to write down a set of differential equations with noise to describe how each bird's opinion changes over time and then to analyze the distribution of these opinions after the dynamics have reached equilibrium [12, 11]. The second is to write down an explicitly probability distribution of the opinions at equilibrium and to analyze this distribution. WIth either approach, we consider a vector of opinions v where v_i a a scalar that represents the ith's bird opinion about an unspecified behavior, for example where or how fast to fly. Both approaches rely on a network of connections between the birds, which we will describe first. Let A be the adjacency matrix such that

$$A_{ij} = \begin{cases} 1 & \text{if } i \text{ pays attention to } j \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

If k_i is the number of neighbors node i has, then $k_i = \sum_k A_{ik}$. We assume that having more neighbors means paying less attention to each so we normalize A to find $\bar{A}$: $\bar{A}_{ij} = \frac{A_{ij}}{k_i}$.

The Laplacian of this matrix is given by $L = \bar{A} - I$, which is symmetric only if A is. We can symmetrize the interactions by defining $S = \frac{1}{2}(A + A^T)$ and P such that $P_{ij} = S_{ij}$ for $j \neq i$ and $P_{ii} = -\sum_k S_{ik}$. Then L and P share the properties that $L\vec{1} = P\vec{1} = \vec{0}$ and that each is negative semi-definite, with the difference being that P is necessarily symmetric. It is more common to use L in a dynamics system approach, so that

$$\dot{v}_i = \frac{1}{k_i} \sum_i (v_j - v_i) + \sqrt{k_i}\xi_i \tag{2}$$

$$\dot{v} = Lv + D\xi \tag{3}$$

where D is a diagonal matrix with entires $\{\sqrt{k_i}\}$ and $E[\xi(t)] = \vec{0}$, $E[\xi(t)\xi(s)^T] = \frac{1}{2}\delta(t-s)I$, and $E[v(0)\xi(t)^T] = \vec{0}$. This describes a system in which each bird changes averages the differences between its own opinions and those of its neighbors and changes its opinion to minimize these differences. Our choice of D assumes that birds with more neighbors make noisier estimates of how to change their opinions.

The Boltzmann distribution in statistical physics is a standard way in which to write an equilibrium distribution. In physics, those states with the highest energy are least likely. In our system, those sets of opinions that are least in consensus will be less likely. Specifically, we will use the following expression to represent the "energy" of an opinion vector

$$H(v) = -v^T P v \tag{4}$$

and the Boltzmann distribution

$$P(v) = \frac{1}{Z} \exp\left(-\frac{1}{2}H(v)\right) \tag{5}$$

where $Z = \int_{\mathbb{R}}^N \exp\left(-\frac{1}{2}H(v)\right) dv$.

With the dynamical approach, in the absence of noise, the system would reach consensus with each $v_i = \alpha$ for some $\alpha \in \mathbb{R}$. Similarly, with the statistical physics approach, consensus states with each $v_i = \alpha$ are most likely. However, all consensus values α are equally likely. It is therefore difficult to describe correlations between the bird's opinions at equilibrium. It becomes easier, however, to consider deviations from consensus. If all v_i are equal, then $v = \alpha\vec{1}$. To consider deviations from this consensus state, it is useful to define a matrix $Q \in \mathbb{R}^{N-1} \times \mathbb{R}^N$ such that each row of Q is orthogonal to $\vec{1}$ and the rows are pairwise orthogonal. Young et al. show in [12, 11] that this is equivalent to a Q such that $QQ^T = I_{N-1}$ and $Q^T Q = I_N - \frac{1}{N}\vec{1}\vec{1}^T$. To simply consider absolute deviations from the average opinion of the flock, we can define $z = v - \langle v \rangle\vec{1} = Q^T Q v$, where $\langle v \rangle = \frac{v_i}{N}$. To consider the component of the opinion vector that is orthogonal to consensus, we can define $y = Qv = Qz$. Let $\Sigma_z = \lim_{t\to\infty} E[zz^T]$ and $\Sigma_y = \lim_{t\to\infty} E[yy^T]$ using the dynamical approach and $C_z = E[zz^T]$ and $C_y = E[yy^T]$ using the statistical physics approach. Since $y = Qz$ and $z = Q^T y$, in either case, we can translate easily between Σ_z and Σ_y or between C_z and C_y. In the Appendix, we derive equations for each of these covariance matrices and summarize the expression for the y covariance matrices in Table 2. The first thing to note is that, if instead of using L we use P in the dynamical approach and all birds have the same number of neighbors, then $C_y = 2\Sigma_y$, as shown in Claim 5. There is therefore a close correspondence between the two approaches.

These covariance matrices allow us to define two group properties. The first is $\mathscr{H}_2$ robustness, as used in [12, 11]. This is a measure of the robustness of the consensus state in which all birds have the same opinion to noise at equilibrium. Specifically,

$$\rho = \frac{1}{\lim_{t\to\infty} E[||y||]} = \frac{1}{\lim_{t\to\infty} \operatorname{Tr}(\Sigma_y)}. \tag{6}$$

The second property is the correlation length of the flock. This is the distance over which birds' opinions tend to be positively correlated with each other. If we know the distances between the birds and the

covariance matrix C_z, we can plot the correlations $\{E[z_i z_j]\}$ between each pair of birds' deviations from consensus against the distance between the birds and then identify the correlation length λ by finding the distance at which these correlations become 0.

So far, we have considered birds whose opinions only depend on the opinions of the rest of the flock. We can additionally incorporate environmental information by assuming that there is an environmental signal such that more informed birds have opinions closer to 1. If we let B be a matrix with $B_i i = 1$ if i can perceive the signal and 0 otherwise, then the dynamical system becomes

$$\dot{v} = Lv + B(\vec{1} - v) + D\xi = L_\mathrm{s}v + B\vec{1} + D\xi \tag{7}$$

where $L_\mathrm{s} = L - B$. Similarly, the statistical physics model becomes

$$P(v) = \frac{1}{Z} \exp\left(-\frac{1}{2}(v - \vec{1})^T(-P + B)(v - \vec{1})\right) = \frac{1}{Z} \exp\left(\frac{1}{2}(v - \vec{1})^T(-P_\mathrm{s})(v - \vec{1})\right) \tag{8}$$

where $P_\mathrm{s} = P - B$. Again, we can find the covariance matrices Σ_z, Σ_y, C_z, and C_y. As was the case without a signal, if we use P in the dynamical approach and all birds have the same number of neighbors, then we find that $C_y = 2\Sigma_y$ (Claim 5). We define ρ and λ as above using the matrices Σ_y and C_z in the presence of a signal.

2.2   Individual-level fitness To find the optimal strategies from the individuals' perspectives, we need a measure of individual-level fitness. We use the dynamical system to describe how well each bird learns about the environmental signal. For a given matrix L, the expected opinion q_i of each individual satisfies the equations

$$\dot{q} = L_\mathrm{s}q + B\vec{1}. \tag{9}$$

To generate a matrix L, we distribute the individuals randomly in space (specifically, in a unit square). has a strategy k_i that dictates how many neighbors it has, so that $k_i = \sum_{j \neq i} \mathbf{I}(A_{ij} \neq 0)$. Each individual then pays attention to its k_i nearest neighbors. One individual is chosen to be a receiver; it and all those individuals within a radius r of the receiver perceive, so that $B_{ii} = 1$ for all those birds within this radius. We do this 1000 times and find $q(1)$ for each matrix, which we refer to as a signaling event. This is meant to simulate a situation in which such that information is present only intermittently and otherwise the birds are moving randomly through space and restructuring their network. If, instead of paying attention to the closest neighbors, a bird choose k_i neighbors at random out of N birds, N_I of whom are informed, the number of informed neighbors follows a hypergeometric distribution:

$$P(j \text{ informed neighbors}) = \frac{\binom{N_\mathrm{I}}{j}\binom{N - N_\mathrm{I}}{k_i - j}}{\binom{N}{k_i}}. \tag{10}$$

This distribution is not an exact description of the probability of j informed neighbors in our model since it ignores spatial proximity, but it is a useful approximation for considering the advantages and disadvantages of each strategy k_i. NOTE: I'M RUNNING AN ANALYSIS WHERE FITNESS IS BASED SOLELY ON THIS DISTRIBUTION, NOT ON THE FULL SPATIAL STRUCTURE OF THE MODEL. I SHOULD HAVE RESULTS SOON. I EXPECT THEM TO BE SIMILAR TO THE FULL MODEL, BUT I WANTED TO MAKE SURE AND TO USE IT AS A WAY OF EXPLAINING THE FULL MODEL RESULTS.

We consider signals that provide information about two types of content: predators and food. If the signal is information that a predator is present, we assume that whichever individual is least aware of the signal will be predated, i.e. the individual i such that q_i is lowest. Then, fitness is proportional to the probability of surviving, i.e. 1 minus the probability of being the least informed over many signaling events. If the signal is information that resources are available, we consider two levels of resource abundance. If there are scarce resources, we assume that whichever individual is most aware of the signal will get access to the resource, i.e. the individual i such that q_i is highest. In this case, fitness is proportional to the probability of getting the resource, i.e. the probability of being the most informed over many signaling events. If there are abundant resources, we assume that the first four individuals to the resources can share it, so that fitness is proportional to the probability of being one

of the four most informed birds. Finally, we consider a combination of predation and resources: fitness is proportional to the 1 minus the probability of being predated times the probability of being the most informed. The behavior that the birds can change to improve their fitness is the number of other birds they pay attention to, k_i. In particular, each individual has a strategy k_i that dictates how many neighbors it has, so that $k_i = \sum_{j \neq i} \mathbf{I}(A_{ij} \neq 0)$.

2.3 Optimization methods To understand what strategies we might expect to find, we are interested in identifying the optimal strategies. We does this in two ways. First, using the framework of adaptive dynamics in finite populations, we identify the evolutionarily stable strategy. This framework assumes that there is a homogeneous population into which a mutant individual tries to invade. If strategies change on a learning rather than an evolutionary timescale, or if there is just a lot of variation in the population, we might expect individuals to try to optimize their strategies in the context of a heterogeneous population. This is our second method of optimization. Given a random set of initial strategies over the group, we allow each bird to choose the strategy that would be best given the rest of the strategies being used and 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 optimization process.

3 Results

3.1 Robustness and correlation length are positively related. In a homogeneous flock, where every bird is using the same strategy, as both the strategy and the radius of the signal change, both $\mathscr{H}_2$ and the correlation length are affected. Figure 1 shows that both $\mathscr{H}_2$ and correlation length are maximized at intermediate strategies when the radius is low and at low strategies when the radius is high. We prove in the Appendix (Appendix Claim 5) that if we use a symmetric matrix to calculate ρ and all birds have the same number of neighbors n, then the covariance matrices used to calculate ρ and λ are identical up to a constant, which suggests that the two properties cannot be optimized independently.

3.2 The ESS number of neighbors depend on the content of the signal. Those birds that can perceive the environmental signal will always have better information than those that cannot. Selection, therefore, acts to improve a bird's ability to learn about the signal when it cannot perceive it. In Appendix Claim 1 we show that an uninformed bird's opinion can be approximated by $\sum_{\text{informed } j} L_{ij}$, that is the total weight given to informed birds. Being the most well-informed uninformed bird, therefore, amounts to having the highest proportion of informed neighbors and being the least well-informed amounts to having the smallest proportion of informed neighbors. As described above, the hypergeometric distribution is a good approximation for the probability distribution of the number of informed neighbors a bird with strategy k_i will have. As shown in Figure 4, under the hypergeometric distribution, having many neighbors makes it unlikely that few of those neighbors will be uninformed, but it also makes it unlikely that all of them will be uninformed (Figure 4). Consequently, a bird with more neighbors will be less likely to be the least informed, but will also be less likely to be the best informed.

These observations explain the ESS number of neighbors we find, which depends on which selection pressure is being applied. Figure 2 shows that the ESS number of neighbors when selection is due to access to scarce resources is always 1. The ESS number of neighbors when selection is due to both access to resources and predation is also always 1. On the other hand, the ESS number of neighbors when selection is due to predation is always high, although not necessarily as high as possible. There is no ESS when selection is due to access to abundant resources.

3.3 The predation ESS increases as the signal becomes more public. The other feature of the ESS number of neighbors when selection is due to predation is that it increases as the signal becomes more public (Figure 2). When trying to avoid begin the least well-informed, having too few neighbors is always disadvantageous because a bird with too few neighbors runs the risk of having too few or

no informed neighbors. When the signal is not widely available, having too many neighbors is also disadvantageous because, even if a bird paying attention to many neighbors pays attention to all the informed birds, the total weight given to those informed birds will be lower than a bird with fewer total neighbors. As more can perceive the signal, the costs of paying attention to uninformed birds disappear and the ESS is to pay attention to all other birds in the flock.

3.4   The birds construct an optimal network when they learn strategies to optimize access to abundant resources. Learned strategies when selection is due to predation are high and learned strategies when selection is due to access to scarce resources or both resources and predation are low (Figure 3). Whereas there is no ESS when selection is due to access to abundant resources, when we let a heterogenous flock learn strategies, we find that most birds in the flock decrease their strategies until they have three or four neighbors under this selection regime. As the birds choose new strategies, they change the average properties of the networks they construct. When they choose strategies according to predation, access to scarce resources, or both, both ρ and λ decrease or, at best, remain constant. However, when they choose strategies to get access to abundant resources, both ρ and λ increase (Figure 3).

4   Discussion

5 Appendix

5.1 Linear algebra claims and theorems

Claim 1 Suppose. $v(t)$ satisfies $\dot{v} = Av + f$ with $v(0) = v_0$, where A has no 0 eigenvalues. If V is the matrix of eigenvectors of A and Λ is diagonal with the eigenvalues of A, then

$$v(t) = V \exp(t\Lambda)V^{-1}v_0 + V\Lambda^{-1}(\exp(t\Lambda) - I)V^{-1}f. \tag{11}$$

We can approximate $v(t) \approx V \exp(t\Lambda)V^{-1}v_0 + V(tI + \frac{t^2}{2}A)V^{-1}f$ if the eigenvalues are not too far from 0. In the case that $A = L_f$ and $f = B\vec{1}$, this suggests that for i that cannot perceive the signal, $v_i(t) \approx \frac{t^2}{2}\sum_{\text{informed } j} L_{ij}$. This approximation is reasonable for i such that $B_{ii} = 0$, but worse for i such that $B_{ii} \neq 0$.

Proof. First we show that (11) solves the differential equation. Note that if $A = V\Lambda V^{-1}$.

$$\begin{aligned}
\dot{v} &= V\Lambda \exp(t\Lambda)V^{-1}v_0 + V\Lambda^{-1}(\Lambda \exp(t\Lambda))V^{-1}f \\
&= V\Lambda V^{-1}V \exp(t\Lambda)V^{-1}v_0 + V \exp(t\Lambda)V^{-1}f \\
&= AV \exp(t\Lambda)V^{-1}v_0 + V \exp(t\Lambda)V^{-1}f - VV^{-1}f + VV^{-1}f \\
&= AV \exp(t\Lambda)V^{-1}v_0 + V(\exp(t\Lambda) - I)V^{-1}f + f \\
&= AV \exp(t\Lambda)V^{-1}v_0 + V\Lambda V^{-1}V\Lambda^{-1}(\exp(t\Lambda) - I)V^{-1}f + f \\
&= AV \exp(t\Lambda)V^{-1}v_0 + A\Lambda^{-1}(\exp(t\Lambda) - I)V^{-1}f + f \\
&= Av(t) + f.
\end{aligned}$$

If $g(x) = \frac{1}{x}(\exp(ax) - 1)$, then $g(x) \approx a + \frac{a^2}{2}x$ for x close to 0. Then, if the eigenvalues of A are not too far from 0, $\Lambda^{-1}(\exp(t\Lambda) - I) \approx tI + \frac{t^2}{2}\Lambda$. If $A = L_f$, then all eigenvalues of A are negative. If N_I is the number of informed individuals, i.e. i such that $B_{ii} \neq 0$, then there are N_I eigenvalues of A close to -2 such that, if k is the index of one of these eigenvalues, V_{ik} is large if and only if i is informed. The rest of the $N - N_I$ eigenvalues are between -2 and 0 and, if k is the index of one of these eigenvalues, V_{ik} is large if and only if i is uninformed. Therefore, the eigenvalue approximation is worst for the informed individuals and best for the uninformed individuals. Using the eigenvalue approximation, $v(t)$ becomes

$$v(t) = V \exp(t\Lambda)V^{-1}v_0 + tVV^{-1}B\vec{1} + \frac{t^2}{2}V\Lambda V^{-1}B\vec{1} = V \exp(t\Lambda)V^{-1}v_0 + tB\vec{1} + \frac{t^2}{2}L_fB\vec{1}$$

Ignoring the informed individuals and the initial conditions, this gives $v_i(t) \approx \sum_{\text{informed } j} L_{ij}$. ∎

Theorem 2 Given any positive symmetric matrix $Q > 0$ there exists a unique positive symmetric] matrix $\Sigma > 0$ satisfyying $A\Sigma + \Sigma A^T = Q$ if and only if the system $\dot{x} = Ax$ is globally asymptotically stable.

Proof. Proof available elsewhere. ∎

Claim 3 Consider the dynamical system,

$$\dot{y} = Ay + v + Q\xi \tag{12}$$

with $E[\xi(t)] = 0$, $E[\xi(t)\xi(s)^T] = \frac{1}{2}\delta(t - s)I$, and $E[y(0)\xi^T(t)] = 0$. If all eigenvalues of A are negative and if $\Sigma_t = E[y(t)y(t)^T]$ and $\Sigma = \lim_{t\to\infty} \Sigma_t$, then Σ satisfies the Lyapunov equation

$$A\Sigma + \Sigma A^T = A^{-1}vv^T + vv^T(A^{-1})^T - QQ^T.$$

Further, if $C_{ij} = \lim_{t\to\infty} E[(y_i(t) - E(y_i(t)))(y_j(t) - E(y_j(t)))]$ then

$$AC + CA^T = -QQ^T$$

Proof.

$$\Sigma_t = E[y(t)y(t)^T]$$
$$\Rightarrow \dot{\Sigma} = E[\dot{y}(t)y(t)^T + y(t)\dot{y}(t)^T]$$
$$= E[(Ay + v + Q\xi)y^T + y(y^T A^T + v^T + \xi^T Q^T)]$$
$$= AE[yy^T] + E[yy^T]A^T + vE[y^T] + E[y]v^T + QE[\xi y^T] + E[y\xi^T]Q^T$$
$$= A\Sigma_t + \Sigma_t A^T + vE[y^T] + E[y]v^T + QE[\xi y^T] + E[y\xi^T]Q^T.$$

Note that we can write a solution to (12) as

$$y(t) = \exp(At)y(0) + \int_0^t \exp(A(t-s))vds + \int_0^t \exp(A(t-s))Q\xi(s)ds. \tag{13}$$

Eq. (13) allows us to write

$$E[y(t)] = \exp(At)y(0) + \int_0^t \exp(A(t-s))vds$$
$$= \exp(At)y(0) - A^{-1}\exp(A(t-s))v|_{s=0}^{s=t} \text{ which exists because } A \text{ has no 0 eigenvalues}$$
$$= \exp(At)y(0) - A^{-1}(I - \exp(At))v.$$

Note that if we take the limit as t goes to infinity,

$$\lim_{t\to\infty} E[(y(t))] = \lim_{t\to\infty} \exp(At)y(0) - A^{-1}(I - \exp(At))v = -A^{-1}v$$

since all eigenvalues of A are negative. Eq. (13) also allows us to write

$$E[y\xi^T] = \exp(At)y(0)E[\xi(t)^T] + \int_0^t \exp(A(t-s))BE[\xi(t)^T]ds + \int_0^t \exp(A(t-s))QE[\xi(s)\xi(t)]ds$$
$$= 0 + 0 + \frac{1}{2}Q$$
$$\Rightarrow \dot{\Sigma} = A\Sigma_t + \Sigma_t A^T + vE[y(t)^T] + E[y(t)]v^T + \frac{1}{2}QQ^T + \frac{1}{2}QQ^T$$
$$= A\Sigma_t + \Sigma_t A^T + vE[y(t)^T] + E[y(t)]v^T + QQ^T$$
$$\Rightarrow \lim_{t\to\infty} \dot{\Sigma} = A\Sigma + \Sigma A^T - vv^T(A^{-1})^T - A^{-1}vv^T + QQ^T$$
$$\Rightarrow A\Sigma + \Sigma A^T = vv^T(A^{-1})^T + A^{-1}vv^T - QQ^T \text{ since } \dot{\Sigma} = 0 \text{ at equilibrium.}$$

Further, by definition of C,

$$C = \Sigma - \lim_{t\to\infty} E[y(t)]E[y(t)^T]$$
$$= \Sigma - (-A^{-1}v)(-A^{-1}v)^T$$
$$= \Sigma - A^{-1}vv^T(A^{-1})^T$$
$$\Rightarrow AC + CA^T = A\Sigma + \Sigma A^T - AA^{-1}vv^T(A^{-1})^T - A^{-1}vv^T(A^{-1})^T A^T$$
$$= A\Sigma + \Sigma A^T - vv^T(A^{-1})^T - A^{-1}vv^T$$
$$= -QQ^T.$$

∎

Fact 4 If $V = \sum_i v_i/N$ and $C = (E[(v_i - V)(v_j - V)])_{ij}$, then $\sum_{ij} C_{ij} = 0$.

Proof.

$$\sum_{ij} E[(v_i - V)(v_j - V)] = \sum_{ij} (E[v_i v_j] - E[v_i V] - E[v_j V] + E[VV])$$

$$= \sum_{ij} E[v_i v_j] - N \sum_i E[v_i V] - N \sum_j E[v_j V] + N^2 E[VV]$$

$$= \sum_{ij} E[v_i v_j] - \sum_{ij} E[v_i v_j] - \sum_{ij} E[v_i v_j] + \sum_{ij} E[v_i v_j]$$

$$= 0$$

$\blacksquare$

In the following sections of the Appendix, transforming an opinion vector x will be useful. To do so, we define a matrix $Q \in \mathbb{R}^{N-1} \times \mathbb{R}^N$ such that each row of Q is orthogonal to $\vec{1}$ and the rows are pairwise orthogonal. Young et al. show in [12] show that this equivalent to finding Q such that $Q\vec{1} = \vec{0}$, $QQ^T = I_{N-1}$, and $Q^T Q = I_N - \frac{1}{N}\vec{1}_N\vec{1}_N^T$. We next define a rotation matrix $R \in \mathbb{R}^N \times \mathbb{R}^N$ such that the last row of R is $\frac{1}{\sqrt{N}}\vec{1}^T$ and the other rows are given by the rows of Q. Since the rows of R are orthonormal, $RR^T = I$ and R is invertible, which implies that $R^T = R^{-1}$. If $z = Q^T Q x$, then $z_i = v_i - \frac{1}{N}\sum_i v_i$. Similarly, if $\bar{y} = Rx$, then $\bar{y}_N = \frac{1}{\sqrt{N}}\sum_i x_i$ and $y = (\bar{y}_1, \ldots, \bar{y}_{N-1})^T = Qx$ gives the component of v that is orthogonal to the consensus vector $\vec{1}$. We can also translate between y and z by noting that $z = Q^T y$ and $y = QQ^T z$ or y is given by the first $N-1$ coordinates of Rz.

$$\boxed{V = \frac{1}{N}\sum_i x_i, \ z = Q^T Q x = x - V\vec{1}, \ \bar{y} = Rx, \ y = Qx = \bar{y}_{i=1}^{N-1}} \tag{14}$$

5.2 Opinion dynamics with noise In our model, each bird in the flock as an opinion x_i, giving an opinion vector x. The $\mathcal{H}_2$ robustness of the opinion dynamics is defined by the inverse of the expected distance at equilibrium from the opinion vector to a consensus state in which all birds have the same opinion, i.e. $x_i = \alpha$ for all i and for some $\alpha \in \mathbb{R}$. That is, the dynamics are robust if, at equilibrium, the opinions are close to being in consensus. We can find this distance by rotating the opinion vector into the space orthogonal to the consensus vector, $\vec{1}$. If $y = Qx$, then $||y|| = y^T y$ describes how far away from consensus x is. Specifically, robustness as in [12] is given by

$$r = \lim_{t \to \infty} \frac{1}{E[||y||]} = \lim_{t \to \infty} \frac{1}{\text{Tr}E[yy^T]}. \tag{15}$$

We also define a full rotation matrix $R \in \mathbb{R}^N \times \mathbb{R}^N$ such that $R = \begin{pmatrix} Q \\ \frac{1}{\sqrt{N}}\vec{1}^T \end{pmatrix}$.

Young et al. [12] model opinion dynamics by assuming that each bird averages the differences between its opinion and each of the opinions of its neighbors and changes its opinion to minimize this difference and that each bird's opinion is subject to independent noise with variance proportional to the number of neighbors each bird has. The equations describing these dynamics can be written as

$$\dot{x}_i = \frac{\sum_{\text{neighbors } j \text{ of } i}(x_j - x_i)}{k_i} + \sqrt{k_i}\xi_i = \sum_j \bar{A}_{ij}x_j - x_i + \sqrt{k_i}\xi_i = \sum_j L_{ij}x_j + \sqrt{k_i}\xi_i. \tag{16}$$

In vector notation this becomes

$$\dot{x} = Lx + D\xi. \tag{17}$$

The equilibrium of these dynamics is a consensus state in which all birds have the some opinion, $x = \alpha\vec{1}$ for some $\alpha \in \mathbb{R}$. In the absence of any environmental signal, the consensus opinion can be any value α. In the presence of an environmental signal, some birds in the flock perceive the signal and change

their opinions to be closer to 1. The environmental forcing is described by a diagonal matrix B where $B_{ii} = b$ if i perceives the environmental signal and $B_{ii} = 0$ otherwise and the dynamics then become

$$\dot{x} = Lx + B(\vec{1} - x) + D\xi = L_\mathrm{s}x + v + D\xi. \tag{18}$$

where $L_\mathrm{s} = L - B$ and $v = B\vec{1}$. If $b = 0$, then $L_\mathrm{s} = L$ and $v = \vec{0}$, so the case with the signal reduces to the case without the signal.

If $\bar{y} = Rx$, then the first $N-1$ coordinates of $\bar{y}$ are $y = Qx$ and the last is $\sum_i x_i / \sqrt{N}$. The dynamics of $\bar{y}$ are described by

$$\begin{aligned}
\dot{\bar{y}} &= R\dot{x} \\
&= RL_\mathrm{s}x + Rv + RD\xi. \\
&= RL_\mathrm{s}R^T\bar{y} + Rv + RD\xi \\
&= \tilde{L}_\mathrm{s}\bar{y} + \tilde{v} + \tilde{D}\xi
\end{aligned}$$

where $\tilde{L}_\mathrm{s} = RL_\mathrm{s}R^T$, $\tilde{v} = Rv = RB\vec{1}$, and $\tilde{D} = RD$. If $b \neq 0$, L_s and therefore $\tilde{L}_\mathrm{s}$ have all negative eigenvalues and $\bar{y} = (0, 0, \ldots 0, \sqrt{N})^T$ is a globally asymptotically stable equilibrium. Therefore, by Theorem 2 and Claim 3, $\lim_{t\to\infty} E[\bar{y}(t)\bar{y}(t)^T] = \Sigma_{\bar{y}}$ where $\tilde{L}_\mathrm{s}\Sigma_{\bar{y}} + \Sigma_{\bar{y}}\tilde{L}_\mathrm{s}^T = -RD^2R^T$. We can find $\Sigma_y = \lim_{t\to\infty} E[y(t)y(t)^T]$ by removing the last row and column of $\Sigma_{\bar{y}}$.

If $b = 0$, L has a 0 eigenvalue corresponding to the eigenvector $\vec{1}$. Therefore, RLR^T also has a 0 eigenvalue, so we can therefore no longer find the desired $\Sigma_{\bar{y}}$. Since the last row of $R = \frac{1}{\sqrt{N}}\vec{1}$, all entries of the the last column of LR^T are 0. Therefore, all entries of the last column of RLR^T are 0. In other words, the dynamics of the first $N-1$ elements of $\bar{y}$, i.e. y, do not depend on the last. This lets us write

$$\dot{y} = QLQ^Ty + Q\xi = \bar{L} + Q\xi$$

where $\bar{L} = QLQ^T$. Another way to see this is to write

$$\begin{aligned}
\dot{y} &= Q\dot{x} \\
&= QLx + Q\xi \\
&= QL(x - \langle x\rangle\vec{1}) + Q\xi \text{ since } L\vec{1} = \vec{0} \\
&= QLQ^TQx + Q\xi \\
&= \bar{L}y + Q\xi.
\end{aligned}$$

Again, by Claim 3 and Theorem 2, we can now find $\lim_{t\to\infty} E[y(t)y(t)^T] = \Sigma_y$ by finding a solution to $\bar{L}\Sigma_y + \Sigma_y\bar{L} = -QD^2Q^T$.

To summarize, in the absence of an environmental signal, Σ_y satisfies $\bar{L}\Sigma_y + \Sigma_y\bar{L}^T = -QD^2Q^T$ where $\bar{L} = QLQ^T$ and in the presence of an environmental signal, Σ_y can be found in the first $N-1$ rows and columns of $\Sigma_{\bar{y}}$ where $\Sigma_{\bar{y}}$ satisfies $\tilde{L}_\mathrm{s}\Sigma_{\bar{y}} + \Sigma_{\bar{y}}\tilde{L}_\mathrm{s}^T = -RD^2R^T$ where $\tilde{L}_\mathrm{s} = RL_\mathrm{s}R^T$.

5.3 Probability distribution at equilibrium A quite different way of modeling social interactions the method used by Bialek et al. in [3], which is to assume the opinions come to equilibrium and write down an expression for the probability of finding the opinion vector in every state. In statistical physics, a Hamiltonian describes the energy of a given state, $H(v)$, and the Boltzmann distribution describes a probability distribution in which the most energetic states are least likely. Analogously, we can write a function $H(v)$ that gives a high energy when the opinions are not in consensus and a low energy when the opinions are in consensus and a probability distribution of opinion vectors that favors less energetic states. In order to do this, we must use a symmetric interaction matrix, rather than the possibly asymmetric $\bar{A}$ we used above. Bialek et al. study a model in which opinions are a three-dimensional velocity vector and the birds can weight the opinions of their nearest neighbors with any value J. In order to simplify this model and make it directly comparable to the opinion dynamics model we described above, we only consider a scalar-valued opinion x_i and we impose the condition

that the weights given to the neighbors' opinions are inversely proportional to the number of neighbors. In particular, we define $S = \frac{1}{2}(A + A^T)$ and P such that $P_{ij} = S_{ij}$ for $j \neq i$ and $P_{ii} = -\sum_k S_{ik}$. This allows us to write down a Hamiltonian

$$H(v) = -v^T P v \tag{19}$$

where the negative sign makes opinions far from consensus give high energies and conversely. The probability distribution over opinion vectors v then becomes

$$P(v) = \frac{1}{Z} \exp\left(-\frac{1}{2}v^T(-P)v\right), \tag{20}$$

where $Z = \int_{\mathbb{R}}^N \exp\left(-\frac{1}{2}H(v)\right) dv$. We can now use this probability distribution to find the covariances and correlations between the birds' fluctuations around consensus. If we let $V = \sum_i v_i/N = \langle v \rangle$ and $z = v - V\vec{1}$, z represents the deviations in each bird's opinion from the consensus opinion. We are interested in finding $C_z = (E[z_i z_j])_{ij}$. To do this, we first find the probability distribution over (z, V):

$$P(z, V) = \frac{1}{Z} \exp\left(-\frac{1}{2}(z + V\vec{1})^T(-P)(z + V\vec{1})\right)$$

$$= \frac{1}{Z} \exp\left(-\frac{1}{2}z^T(-P)z\right) \text{ since } P\vec{1} = P^T\vec{1} = \vec{0}.$$

Since P has a 0 eigenvalue, it is not invertible. However, z is orthogonal to the eigenvector of P with eigenvalue 0. Therefore, the covariance of z is $C_z = -P^\dagger$. By dividing each entry C_{zij} of the covariance matrix by $C_{zii}C_{zjj}$, the covariance matrix can be used to find the correlation matrix.

The Hamiltonian framework, like the opinion dynamics framework, can also accommodate an environmental signal. In this case, the appropriate Hamiltonian is

$$H(v) \quad = -(v - \vec{1})^T P_s (v - \vec{1}) \tag{21}$$

where $P_s = P - B$. It is clear that H is minimized when $v = \vec{1}$. Again, we are interested in the deviations from the consensus opinion. As above, let $V = \sum_i v_i/N$ and $z = v - V\vec{1}$ so that $v = z + V\vec{1}$.

$$P(v) = \frac{1}{Z} \exp\left(-\frac{1}{2}(v - \vec{1})^T(-P_s)(v - \vec{1})\right)$$

$$= \frac{1}{Z} \exp\left(-\frac{1}{2}(z - (1 - V)\vec{1})^T(-P_s)(z - (1 - V)\vec{1})\right)$$

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 $\mathbb{R}^{N-1}$ and integrate over all of $\mathbb{R}^{N-1}$. Therefore, we define $y = Rz = RQ^TQv = Qv$. Since $\sum_i z_i = 0$, $y_N = 0$ so V describes the mean of v and y describes the $N-1$ dimensional space of deviations from the consensus opinion.

$$P(v) = \frac{1}{Z} \exp\left(-\frac{1}{2}(R^T y - (1 - V)\sqrt{N}R^T e_N)^T(-P_s)(R^T y - (1 - V)\sqrt{N}R^T e_N)\right)$$

$$= \frac{1}{Z} \exp\left(-\frac{1}{2}(y - (1 - V)\sqrt{N}e_N)^T(-RP_sR^T)(y - (1 - V)\sqrt{N}e_N)\right)$$

$$= \frac{1}{Z} \exp\left(-\frac{1}{2}(y + \sqrt{N}Ve_N - \sqrt{N}e_N)^T\tilde{P}_s(y + \sqrt{N}Ve_N - \sqrt{N}e_N)\right)$$

where $\tilde{P}_s = -RP_sR^T$. To find $P(z)$ we can integrate out V. However, to consider this as a probability density function over y and V, we have to consider the determinant of the transformation from v to y and V. The transformation can be represented by $\bar{R} = \begin{pmatrix} Q \\ 1/\sqrt{N}R_{N.} \end{pmatrix}$, so that $\det(\bar{R}) = \frac{1}{\sqrt{N}}\det(R) = \frac{1}{\sqrt{N}}$.

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

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

Therefore, $E[y_i] = 0$ for all i and $E[yy^T] = (\tilde{P}_s - \frac{1}{\tilde{P}_{sNN}}\tilde{P}_{s\cdot N}\tilde{P}_{sN\cdot})^{-1}$. Then,

$$E[zz^T] = R^T\left(\tilde{P}_s - \frac{1}{\tilde{P}_{sNN}}\tilde{P}_{s\cdot N}\tilde{P}_{sN\cdot}\right)^{-1}R = \left(-P_s - \frac{1}{\tilde{P}_{sNN}N}B\vec{1}\vec{1}^T B\right)^{-1}$$

Let N_I be the number of birds that can perceive the signal, i.e. N_I is the number of non-zero elements along the diagonal of B. Then, $\tilde{P}_{sNN} = \frac{b}{N}N_I$. Finally, we find that

$$C_z = \left(-P_s - \frac{1}{bN_I}B\vec{1}\vec{1}^T B\right)^{-1} \tag{22}$$

If $P' = P_s + \frac{1}{bN_I}B\vec{1}\vec{1}^T B$, then $P'\vec{1} = -B\vec{1} - \frac{1}{bN_I}B\vec{1}\vec{1}^T B\vec{1} = -B\vec{1} - \frac{bN_I}{bN_I}B\vec{1} = \vec{0}$, so we cannot invert $-P'$ to find the covariance matrix. However, as above, we can discard the 0 eigenvector since we know that $z \cdot \vec{1} = 0$. This gives us $C_z = (-P_s - \frac{1}{bN_I}B\vec{1}\vec{1}^T B)^\dagger$.

5.4 Connecting two frameworks Table 2 shows how to find the covariance of y depending on the framework being used and whether or not a signal is present.

Claim 5 If we use P instead of L with the noisy dynamics framework and all birds have the same number of neighbors n, then $C_y = \frac{2}{n}\Sigma_y$.

Proof. If we use P in the noisy dynamics framework, Σ_y satisfies

$$QPQ^T\Sigma_y + \Sigma_y QP^TQ^T = -QD^2Q^T. \tag{23}$$

If all birds have n neighbors, then D^2 has n along the diagonal and 0 elsewhere and $-QD^2Q^T = -nI_{N-1}$. Now consider

$$
\begin{aligned}
QPQ^TC_y + C_yQP^TQ^T &= QPQ^T(-QP^\dagger Q^T) + (-QP^\dagger Q^T)QP^T \\
&= -QP(I_N - \frac{1}{N}\vec{1}\vec{1}^T)P^\dagger Q^T - QP^\dagger(I_N - \frac{1}{N}\vec{1}\vec{1}^T)P^T \\
&= -QPP^\dagger Q^T - QP^\dagger P^TQ^T \text{ since } P\vec{1} = \vec{0} \\
&= -Q(I_N - \frac{1}{N}\vec{1}\vec{1}^T)Q^T - Q(I_N - \frac{1}{N}\vec{1}\vec{1}^T)Q^T \\
&= -QQ^T - QQ^T \text{ since } Q\vec{1} = \vec{0} \\
&= -2I_{N-1}
\end{aligned}
$$

$$\Rightarrow QPQ^T\left(\frac{n}{2}C_y\right) + \left(\frac{n}{2}C_y\right)QP^TQ^T = -nI_{N-1}$$

$$\Rightarrow \frac{n}{2}C_y = \Sigma_y$$

since there is a unique solution to this Lyapunov equation. ∎

A similar proof shows the equivalence of the two covariance matrices in the presence of a signal, as long as P_s is used in the noisy dynamical system and all birds have the same number of neighbors.

References

[1] M. Ballerini, N. Cabibbo, R. Candelier, A. Cavagna, E. Cisbani, I. Giardina, V. Lecomte, A. Orlandi, G. Parisi, A. Procaccini, M. Viale, and V. Zdravkovic. Interaction ruling animal collective behavior depends on topological rather than metric distance: Evidence from a field study. *Proceedings of the National Academy of Science*, 105(4):1232–1237, January 2008.

[2] William Bialek, Andrea Cavagna, Irene Giardina, Thierry Mora, Oliver Pohl, Edmondo Silvestri, Massimiliano Viale, and Aleksandra Walczak. Social interactions dominate speed control in driving natural flocks toward criticality. *arXiv*, 2013.

[3] William Bialek, Andrea Cavagna, Irene Giardina, Thierry Mora, Edmondo Silvestri, Massimiliano Viale, and Aleksandra M. Walczak. Statistical mechanics for natural flocks of birds. *Proceedings of the National Academy of Science*, 109(13):4786–4791, March 2012.

[4] Nikolai W.F. Bode, Daniel W. Franks, and A. Jamie Wood. Limited interactions in flocks: relating model simulations to empirical data. *Journal of The Royal Society Interface*, page rsfi20100397, 2010.

[5] Andrea Cavagna, Alessio Cimarelli, Irene Giardina, Giorgio Parisi, Raffaele Santagati, Fabio Stefanini, and Massimiliano Viale. Scale-free correlations in starling flocks. *Proceedings of the National Academy of Science*, 107(26):11685–11870, 2009.

[6] Laurie Landeau and John Terborgh. Oddity and the 'confusion effect' in predation. *Animal Behaviour*, 34(5):1372–1380, 1986.

[7] Ryan Lukeman, Yue-Zian Li, and Leah Edelstein-Keshet. Inferring individual rules from collective behavior. *Proceedings of the National Academy of Science*, 107(28):12576–12580, 2010.

[8] Yilun Shang and Roland Bouffanais. Influence of the number of topologically interacting neighbors on swarm dynamics. *Scientific reports*, 4, 2014.

[9] Ariana Strandburg-Peshkin, Colin R. Twomey, Nikolai W.F. Bode, Albert B. Kao, Yael Katz, Christos C. Ioannou, Sara B. Rosenthal, Colin J. Torney, Hai Shan Wu, Simon A. Levin, and Iain D. Couzin. Visual sensory networks and effective information transfer in animal groups. *Current Biology*, 23(17):R709–R711, 2013.

[10] Colin J. Torney, Tommaso Lorenzi, Iain D. Couzin, and Simon A. Levin. Social information use and the evolution of unresponsiveness in collective systems. *Journal of The Royal Society Interface*, 12(103):20140893, 2015.

[11] George F. Young, Luca Scardovi, Andrea Cavagna, Irene Giardina, and Naomi E. Leonard. Starling flock networks manage uncertainty in consensus at low cost. *PLOS Computational Biology*, 9, 2013.

[12] George Forrest Young, Luca Scardovi, and Naomi Ehrich Leonard. Robustness of noisy consensus dynamics with directed communication. *Proceedings of the American Control Conference*, 2010.
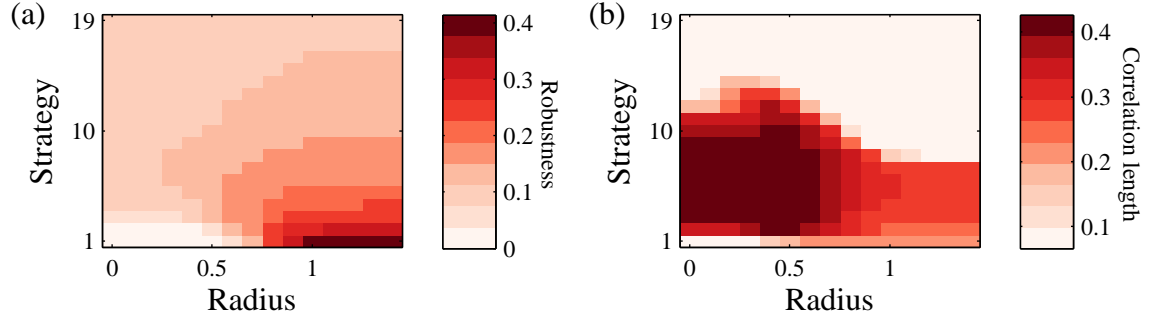
6 Figures

Figure 1: Both $\mathscr{H}_2$ and correlation length are maximized at intermediate strategies when there are few informed birds and at low strategies when there are many informed birds. In (a), we show the average $\mathscr{H}_2$ robustness over 1000 signaling events, where all birds use the same strategy and the signal is available to all birds within a circle of a given radius. In (b), we show the average correlation length over 1000 signaling events, where all birds use the same strategy and the signal is available to all birds within a circle of a given radius.
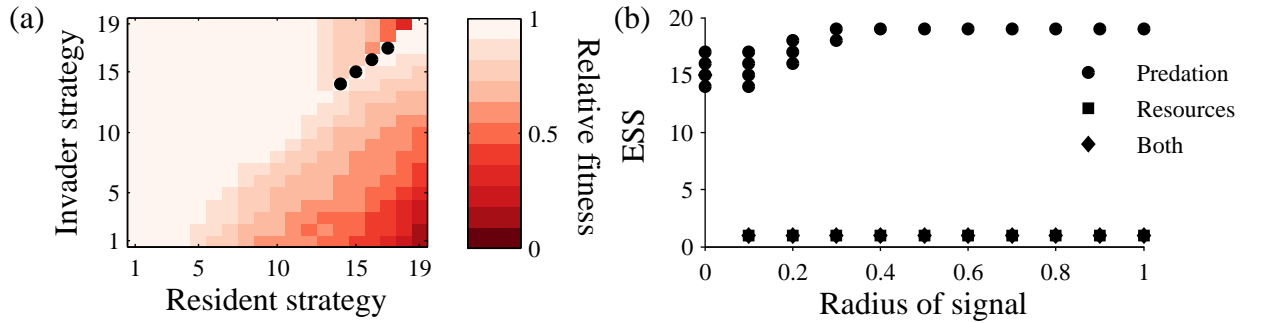

Figure 2: The ESS number of neighbors is always higher if selection is due to predation than due to resources or to both. When selection is due to predation, the (possibly multiple) ESS strategies increase with the radius of the signal. In (a), we show the relative fitness of an invader as a function of the invader and resident strategies when the radius $r = 0.1$, where fitness depends on escaping predation. The circles indicate the ESS strategies, i.e. those resident strategies such that no invader has a higher fitness than the resident. In (b), we show the (possibly multiple) ESS strategies as a function of the radius of the signal. The circles show the ESS when selection is due to predation, the squares when selection is due to access to scarce resources, and the diamonds when selection is due to both. The diamonds and squares overlap. There are no ESS when selection is due to access to abundant resources. Parameters: in both panels $N = 20$, $b = 1$, $T = 1$.

7 Tables

Table 1: Table of matrices and variables used in the text.

A	adjacency matrix	$A_{ij} = \begin{cases} 1 & \text{if } i \text{ pays attention to } j \\ 0 & \text{otherwise} \end{cases}$
$\bar{A}$	normalized adjacency matrix	$\bar{A}_{ij} = \begin{cases} \frac{1}{k_i} & \text{if } i \text{ pays attention to } j \\ 0 & \text{otherwise} \end{cases}$
B	diagonal signal matrix	$B_{ij} = \begin{cases} b & \text{if } i = j \text{ and } i \text{ perceives the environmental signal} \\ 0 & \text{otherwise} \end{cases}$
λ	correlation length	
Λ	diagonal matrix of eigenvalues of P	
L	opinion dynamics w/o signal	$L = \bar{A} - I$
L_{s}	opinion dynamics with signal	$L_{\mathrm{s}} = L - B$
$\bar{L}$	rotated opinion dynamics w/o signal	$\bar{L} = QLQ^T$
$\tilde{L}_{\mathrm{s}}$	rotated opinion dynamics w/ signal	$\tilde{L} = RL_{\mathrm{s}}R^T$
S	symmetrized interaction matrix	$S = \frac{1}{2}(\bar{A} + \bar{A}^T)$
P	energy interactions w/o signal	$P_{ij} = \begin{cases} S_{ij} & \text{for } j \neq i \\ -\sum_k S_{ik} & \text{for } j = i \end{cases}$
P_{s}	energy interactions with signal	$P_{\mathrm{s}} = P - B$
$\tilde{P}_{\mathrm{s}}$		
Q	$N - 1 \times N$ rotation matrix	$Q \in \mathbb{R}^{N-1} \times \mathbb{R}^N \text{such that } Q\vec{1} = \vec{0}, QQ^T = I_{N-1}, Q^TQ = I_N - \frac{1}{N}\vec{1}_N\vec{1}_N^T$
R	full rotation matrix	$\begin{pmatrix} Q \\ \frac{1}{\sqrt{N}}\vec{1}^T \end{pmatrix}$
ρ	$\mathscr{H}_2$ robustness	$\lim_{t\to\infty} \frac{1}{E[||y||]}$ where $y = Qx$
W	matrix of eigenvectors of P	

Table 2: $P^\dagger$ is the inverse of P with the 0 eigenvector / value removed.

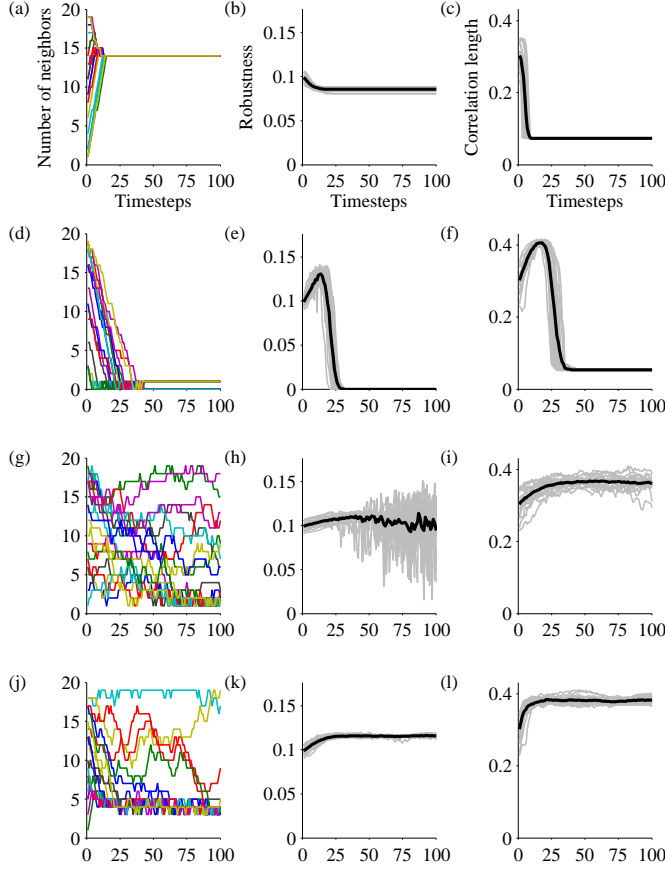|	Without signal	With signal
From noisy dynamics	$\bar{L}\Sigma_y + \Sigma_y\bar{L}^T = -QD^2Q^T$	$\tilde{L}_{\mathrm{s}}\Sigma_{\bar{y}} + \Sigma_{\bar{y}}\tilde{L}_{\mathrm{s}}^T = -RD^2R^T$, $\Sigma_y = (\Sigma_{\bar{y}})_{i,j=1}^{N-1}$
From Boltzmann equilibrium	$C_y = -QP^\dagger Q^T$	$C_y = (\tilde{P}_{\mathrm{s}} - \frac{1}{\tilde{P}_{\mathrm{s}NN}}\tilde{P}_{\mathrm{s}\cdot N}\tilde{P}_{\mathrm{s}N\cdot})^\dagger$
		or $C_y = Q(-P_{\mathrm{s}} - \frac{1}{bN_\mathrm{I}}B\vec{1}\vec{1}^T B)^\dagger Q^T$

Figure 3: When the birds change their number of neighbors to improve their own fitness, they settle on strategies close to the ESS strategies for that selection regimes. When selection is due to predation, they settle on high strategies; when selection is due to access to few resources or to both resources and predation, they settle on low strategies; when selection is due to access to abundant resources, they settle in intermediate strategies. Only under selection due to generous access to resources does this optimization lead to an increase in both $\mathscr{H}_2$ robustness and correlation length. The upper row shows results from selection based to predation, the second row shows results from selection based on access to scarce resources, the third row shows results from selection based on both resources and predation, and the last row shows results from selection based on generous to abundant resources. In each panel, the horizontal axis represents the number of times the birds are allowed to choose optimal strategies. The first column shows one example of how the birds' strategies change over time. The second column shows, for various initial conditions, how the $\mathscr{H}_2$ robustness 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 changes over time, with the black line shows the average of the 20 gray lines. NOTE: I'M RUNNING THIS FOR MORE TIMESTEPS TO SEE IF THE BOTTOM TWO ROWS CONVERGE MORE RELIABLY. Parameters: $N = 20$, $b = 1$, $r = .1$, $T = 1$.
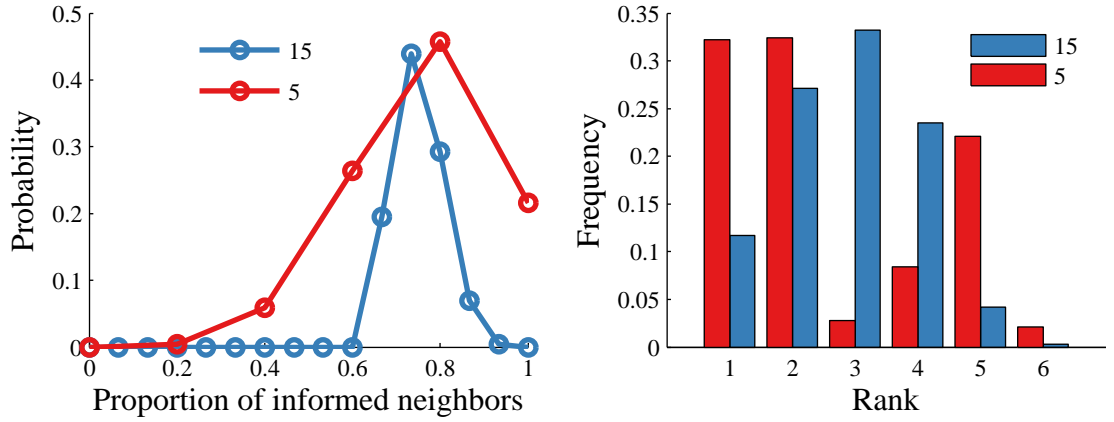
Figure 4: If a fixed number of birds are informed and birds choose neighbors randomly without regard to spatial proximity, the probability of having any number of informed neighbors follows the hypergeometric distribution. A bird with more neighbors is less likely to have a low proportion of informed neighbors, but is also less likely to have a high proportion of informed neighbors. In (a), we show the hypergeometric distribution, giving the probability of having a number of informed neighbors, normalized by the total number of neighbors, for two strategies, 15 in blue and 5 in red. In (b), we consider a group made up of 19 birds using strategy 15 and 1 using strategy 5. For each bird, we draw a number of informed neighbors from the hypergeometric distribution 1000 times and rank the birds according to the proportion of their neighbors that are informed. We show the frequency with which the bird using 5 neighbors and one of the birds using 15 neighbors is ranked 1st through 6th. The bird with 5 neighbors is ranked 1st much more often, but is also ranked 6th more often. Parameters: $N = 20$.