

# Evolution of Information Gathering Strategies

Eleanor Brush  
October 31, 2014

## Contents

<b>1</b>	<b>Abstract</b>	<b>1</b>
<b>2</b>	<b>Model</b>	<b>1</b>
2.1	Learning dynamics . . . . .	1
2.2	Individual-level fitness . . . . .	2
2.3	Group-level performance . . . . .	2
2.4	Optimization methods . . . . .	2
<b>3</b>	<b>Results</b>	<b>2</b>
3.1	ESS strategies . . . . .	2
3.2	Learned strategies . . . . .	4
3.3	Relationship between group properties . . . . .	4
<b>4</b>	<b>Discussion</b>	<b>4</b>
<b>5</b>	<b>Appendix</b>	<b>5</b>
5.1	H2 Norm . . . . .	5
5.2	Calculating correlations . . . . .	5
5.3	Equivalence of dynamics and Hamiltonian . . . . .	6
5.4	Numerically evaluating dynamics . . . . .	7
<b>6</b>	<b>Figures</b>	<b>9</b>

## 1 Abstract

Birds in flocks, like animals in many types of social groups, can use their peers to learn about the environment and update their opinions about where to move in that environment. Previous work on starlings has measured the number of other birds an individual pays attention to and shown that this number of “neighbors” leads to a social network that is conducive to the whole flock reaching consensus. It is unclear, however, why an individual would change its behavior, i.e. the number of neighbors it learns from, to achieve this outcome. Individual birds should optimize how well they learn about the environment. There are (at least) two features of the environment that are important to an individual bird’s fitness—the location of a predator and the location of resources—and it can try to learn about these from its peers. In this work, we identify the optimal strategies for birds under selection pressure dictated by these types of information and explore the relationship between the robustness measure and the correlation length of the flock, a measure commonly used to describe how close a flock’s structure is to criticality.

## 2 Model

**2.1 Learning dynamics** An individual’s opinion changes with a probability that depends on its social interactions and the strength of the environmental stimulus. Specifically, an individual’s opinion  $\sigma_i$  can be either 1 or  $-1$  and  $\sigma_i$  switches with probability  $w_i(\sigma_i) = \frac{1}{2}(-\sigma_i \sum_j A_{ij} \sigma_j)(1 - \beta_i \sigma_i)$  where  $A_{ij}$  indicates the strength of the connection from  $j$  to  $i$  and  $\beta_i$  indicates the strength of the environmental stimulus perceived by  $i$ . The stochastic dynamics of the opinions can be written down with a master

equation [4]. The vector  $\vec{v}(t)$  of expected values,  $q_i(t) = \langle \sigma_i(t) \rangle$ , satisfies the following differential equations: [4]

$$\frac{d\vec{v}}{dt} = M\vec{v} + \vec{\beta}^T(\vec{1} - \vec{v}).$$

We simplify this by writing  $\frac{d\vec{v}}{dt} = A\vec{v} + \vec{\beta}^T\vec{1}$ , where  $A = M - B$  and  $B$  is a diagonal matrix with the values of  $\vec{\beta}$  along the diagonal. In the absence of the environmental signal, there is no trend in the dynamics. In the presence of the environmental signal, however, those individuals who perceive the signal will be more likely to switch to having an opinion of 1 so their expected value will increase, and consequently so will the values of the those paying attention to the informed individuals. Thus, having a higher expected value can be interpreted as having more information about the information and having a lower expected value can be interpreted as having less information.

**2.2 Individual-level fitness** To find the optimal strategies from the individuals' perspectives, we need a measure of individual-level fitness, which we define as follows. Each individual has a strategy that determines how many other individuals he pays attention to,  $n_i = \sum_{j \neq i} \mathbf{I}(A_{ij} \neq 0)$ . We distribute the individuals randomly in space. Each individual then pays attention to its  $n_i$  nearest neighbors. One individual is chosen to be a receiver; it and all those individuals within a radius  $r$  of the receiver perceive an external signal of strength  $\beta$ . We find the expected opinion of each individual  $q_i$  after a period of time  $T$ . We refer to each such set of random positions and receiver as a signaling event.

We consider two types of signals: 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. We therefore find, over many signaling events, the probability that an individual will be eaten and an individual's fitness is given by the probability of surviving, i.e. 1 minus the probability of being predated. If the signal is information that resources are available, 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 being the first to reach the resources.

**2.3 Group-level performance** We consider two group-level properties that indicate how cohesive the group is. The first property is  $\mathcal{H}_2$  robustness, as used in [9, 8]. This is a measure of the robustness of the consensus state in which all birds have the same opinion to noise at equilibrium, as used in (see Appendix Sec. 5.1). The second property is the correlation length of the flock. This is the distance at which the average correlations between birds changes from positive to negative, i.e. the distance over which birds' opinions tend to be positively correlated with each other (see Appendix Sec 5.2).

**2.4 Optimization methods** To understand what strategies we might expect to find, we are interested in identifying the optimal strategies. We do 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 ESS strategies** Surprisingly, even without imposing costs on paying attention to more neighbors, paying attention to the whole flock is not usually the best strategy. The ESS strategy of the number of neighbors to pay attention strongly depends on which selection pressure is being applied. If a focal

bird pays attention to many others, its opinion will be similar to theirs. This will prevent it from being the most informed and hence the first to locate resources. Thus, the uninvadable strategy is to pay attention to few neighbors when selection is due to knowledge about resources. On the other hand, having many neighbors will prevent the focal bird from being the least informed and thus the most likely to be eaten. Thus, the uninvadable strategy is to pay attention to many neighbors when selection is due to knowledge about predators. Figure 1a shows the ESS strategy when selection is due to knowledge about resources is always lower than the ESS strategy when selection is due to knowledge about predators (Figure 1a).

The ESS strategies when selection is due to knowledge about predators show two interesting features. First, there are many parameters for which there are multiple ESS strategies. It might seem that if a strategy outperforms all others when it is the resident strategy, it should also outperform other strategies as an invader. However, when there are multiple ESS strategies, if we take any two of those ESS strategies, each is uninvadable by the other. This indicates how important the social environment set by the resident strategy is. The pairwise invisibility plot in Figure 1b shows that low strategies are always costly and that high strategies are only costly when the resident strategy is high.

If a bird has the environmental signal, it will not be predated, regardless of the number of neighbors it has. The differences in performance only appear when comparing two birds, neither of whom have the signal. To understand the costs of having too few or too many neighbors, we compare the performance of two focal birds, a resident and an invader, neither of whom has the signal. There are two main cases: there can be a single informed bird or multiple informed birds. We further break down the first case depending on which of these two focal birds pays attention to the informed bird: neither, the resident, the invader, or both. For each of these cases, we find the probability that the invader is predated. An invader with one fewer neighbor than the resident always has probability of being predated greater than chance ( $1/N$ ) because the resident often has an informed neighbor while the invader has no or fewer informed neighbors and thus a less informed opinion (Figure 1c). The total probability of being predated stays more or less constant regardless of the resident strategy, with a slight decrease for high resident strategies (Figure 1b and c). An invader with one more neighbor than the resident is only more likely than chance to be predated when the resident strategy is high enough (Figure 1d). This invader performs most poorly when both it and the focal resident pay attention to the informed bird(s). In this case, an invader with one more neighbor has more naive neighbors and thus a less informed opinion. This is increasingly problematic as the resident strategy increases and both the focal birds are likely to pay attention to the informed bird(s).

These opposing forces caused by the costs of not paying attention to informed neighbors and the costs of paying attention to more uninformed neighbors also explains the second feature of the ESS strategies: they are non-monotonic with respect to the radius of the signal, i.e. how public the signal is (Figure 1a). Figure 1a shows that there are essentially three categories of radii: When the radius is low, the signal is private, with only a few birds having true information. When the radius is intermediate, the signal is public, with many birds having true information. When the radius is high, the signal is universal, with essentially all birds in the flock having true information. When the signal is private, paying attention to too few birds is costly because with more neighbors it is more likely to have a truly informed neighbor, but paying attention to too many uninformed birds can also be costly. This leads to a high but not maximal ESS strategy. When the signal is public, the costs of paying attention to uninformed birds are reduced and the ESS strategy increases to become as high as possible. When the signal is universal, not paying attention to informed birds does not impose a cost. Additionally, in a group in which the residents pay attention to all of their neighbors, any mutant strategy can get just as much information by paying attention to fewer neighbors so the maximal strategy is no longer an ESS (Figure 1a).

Another way to explain the non-monotonic ESS number of neighbors is to consider the correlation length of a homogeneous flock as a function of the radius of the signal. If a flock is all using the same strategy, its average correlation length depends much more strongly on the radius of the signal than the strategy being used (Figure 2a). In particular, the correlation length first rises and then falls as the radius is increased. The rise and fall of the correlation length correspond to the rise and fall of the ESS number of neighbors and if we plot the ESS number of neighbors for various radii of the signal against the average correlation length of a homogeneous flock for that radii, we find a positive

relationship (Figure 2b). This is because when correlations in information die off very quickly (i.e. a low correlation length), having more neighbors is costly because birds even slightly away from the informed birds will have poor information. When there are strong correlations in information across the flock, on the other hand, these costs are reduced because every bird in the flock has on average higher quality information.

**3.2 Learned strategies** The equilibrium strategies when the individuals are allowed to learn are similar to the ESS strategies: learned strategies when selection is due to predation are always higher than learned strategies when selection is due to resources (Figure 3). Surprisingly, even when the group starts with heterogeneous strategies, they often reach a homogeneous equilibrium. Under selection due to predation, the  $\mathcal{H}_2$  robustness and correlation length of the flock tend to decrease, whereas under selection due to resources,  $\mathcal{H}_2$  robustness is relatively constant and correlation length tends to increase (Figure 3).

**3.3 Relationship between group properties** Figure 3 shows that there is a relationship between two group properties: the  $\mathcal{H}_2$  norm and the correlation length of the flock.

## 4 Discussion

explore vs. exploit risk aversion vs. risk tolerance

expect single ESS strategies but social information depends on the social environment set by the resident strategy expect ESS strategies to decrease as the signal becomes more public, but we find the opposite

## 5 Appendix

**5.1 H2 Norm** Let  $y$  represent the distance between the vector of opinions and consensus, i.e.  $y = QxQ^T$ , where  $Q \in \mathbb{R}^{n-1 \times n}$  is such that  $Q\vec{1} = 0$ ,  $QQ^T = I_{n-1}$ , and  $Q^T Q = I_n - \frac{1}{n}\vec{1} \times \vec{1}^T$ . Then

$$\dot{y}(t) = -\bar{L}y(t) + Q\xi(t)$$

where  $\bar{L} = QLQ^T$ . As shown in [9], if  $\Sigma_y(t) = \mathbb{E}[y(t)y(t)^T]$ ,  $\Sigma_y$  at equilibrium solves

$$0 = -\bar{L}\Sigma_y - \Sigma_y\bar{L}^T + D. \quad (1)$$

$$\begin{aligned} \mathcal{H}_2 &= \sqrt{\text{Tr}(\Sigma_y)} \\ \mathcal{H}_2 \text{ robustness} &= \frac{1}{\mathcal{H}_2} \end{aligned}$$

**5.2 Calculating correlations** The following is derived from the work in [2].

$$H(\vec{v}) = -\vec{v}^T A \vec{v} - \vec{\beta}^T \vec{v}$$

Define  $V = \sum_k v_k / N$  and  $\vec{y} = (\vec{v} - V\vec{1}) / V$ . Then  $\vec{v} = V(\vec{1} + \vec{y})$ .

$$\begin{aligned} H(\vec{y}, V) &= -V^2(\vec{1}^T + \vec{y}^T)A(\vec{1} + \vec{y}) - \vec{\beta}^T V(\vec{1} + \vec{y}) \\ &= -V^2(\vec{1}^T + \vec{y}^T)A\vec{y} - V\vec{\beta}^T \vec{y} - V\vec{\beta}^T \vec{1} \text{ since } A\vec{1} = \vec{0} \\ &= -V^2\vec{y}^T A \vec{y} + (-V^2\vec{1}^T A - V\vec{\beta}^T)\vec{y} - V\vec{\beta}^T \vec{1} \end{aligned}$$

Define  $Q \in M_{N-1, N}$  that rotates an  $N$ -vector away from the consensus vector and define  $\vec{z} = Q\vec{y}$  so that  $\vec{y} = Q^T \vec{z}$  since  $\sum_k y_k = 0$ .

$$\begin{aligned} H(\vec{z}, V) &= -V^2\vec{z}^T Q A Q^T \vec{z} + (-V^2\vec{1}^T A Q^T - V\vec{\beta}^T Q^T)\vec{z} - V\vec{\beta}^T \vec{1} \\ &= V^2\vec{z}^T P \vec{z} + (-V^2\vec{\sigma}_1^T - V\vec{\sigma}_2^T)\vec{z} - V\vec{\beta}^T \vec{1} \\ \Rightarrow P(\vec{z}, V) &= \frac{1}{\mathcal{Z}} \exp \left( -V^2\vec{z}^T P \vec{z} + (V^2\vec{\sigma}_1^T + V\vec{\sigma}_2^T)\vec{z} + V\vec{\beta}^T \vec{1} \right) \end{aligned}$$

where  $P = -Q A Q^T$ ,  $\vec{\sigma}_1 = Q A^T \vec{1}$ , and  $\vec{\sigma}_2 = Q \vec{\beta}$ .

**Fact 1**

$$\begin{aligned} \int \exp \left( -\frac{1}{2}\vec{z}^T P \vec{z} + \vec{\sigma}^T \vec{z} \right) d^{N-1}z &= \sqrt{\frac{(2\pi)^{N-1}}{\det P}} \exp \left( -\frac{1}{2}\vec{\sigma}^T P^{-1} \vec{\sigma} \right) \\ \Rightarrow \int \exp \left( -\vec{z}^T P \vec{z} + \vec{\sigma}^T \vec{z} \right) d^{N-1}z &= \sqrt{\frac{(2\pi)^{N-1}}{2^{N-1} \det P}} \exp \left( -\frac{1}{4}\vec{\sigma}^T P^{-1} \vec{\sigma} \right) \\ &= \sqrt{\pi^{N-1}} \sqrt{\frac{1}{\det P}} \exp \left( -\frac{1}{4}\vec{\sigma}^T P^{-1} \vec{\sigma} \right) \end{aligned}$$

$$\begin{aligned}
\mathcal{Z} &= \int_{-\infty}^{\infty} \int_{\mathbb{R}^{N-1}} \exp \left( V^2 \vec{z}^T P \vec{z} + (V^2 \vec{\sigma}_1^T + V \vec{\sigma}_2^T) \vec{z} + V \vec{\beta}^T \vec{1} \right) d^N z dV \\
&= \int_{-\infty}^{\infty} \exp(V \vec{\beta}^T \vec{1}) \int_{\mathbb{R}^{N-1}} \exp \left( V^2 \vec{z}^T P \vec{z} + (V^2 \vec{\sigma}_1^T + V \vec{\sigma}_2^T) \vec{z} \right) d^N z dV \\
&= \sqrt{\pi^{N-1}} \int_{-\infty}^{\infty} \exp(V \vec{\beta}^T \vec{1}) \times \left( \sqrt{\frac{1}{V^{2(N-1)} \det P}} \right) \exp \left( -\frac{1}{4} \vec{\sigma}_1^T P^{-1} \vec{\sigma}_1 \right) \exp \left( -\frac{1}{4V} \vec{\sigma}_2^T P^{-1} \vec{\sigma}_2 \right) dV \\
&= \sqrt{\pi^{N-1}} \sqrt{\frac{1}{\det P P}} \exp \left( -\frac{1}{4} \vec{\sigma}_1^T P^{-1} \vec{\sigma}_1 \right) \int_{-\infty}^{\infty} \exp(V \vec{\beta}^T \vec{1}) \times \left( \sqrt{\frac{1}{V^{2(N-1)}}} \right) \exp \left( -\frac{1}{4V} \vec{\sigma}_2^T P^{-1} \vec{\sigma}_2 \right) dV \\
&= \sqrt{\pi^{N-1}} \sqrt{\frac{1}{\det P}} \exp \left( -\frac{1}{4} \vec{\sigma}_1^T P^{-1} \vec{\sigma}_1 \right) \int_{-\infty}^{\infty} \frac{1}{V^{(N-1)}} \exp \left( V \vec{\beta}^T \vec{1} - \frac{1}{4V} \vec{\sigma}_2^T P^{-1} \vec{\sigma}_2 \right) dV \\
&= \sqrt{\pi^{N-1}} \sqrt{\frac{1}{\Pi_a \lambda_a}} \exp \left( -\frac{1}{4} \vec{\sigma}_1^T P^{-1} \vec{\sigma}_1 \right) \int_{-\infty}^{\infty} \frac{1}{V^{(N-1)}} \exp \left( V \vec{\beta}^T \vec{1} - \frac{1}{4V} \vec{\sigma}_2^T P^{-1} \vec{\sigma}_2 \right) dV
\end{aligned}$$

where  $\lambda_a$  are the eigenvalues of  $P$

$$\Rightarrow -\log(\mathcal{Z}) = \text{constants} + \frac{1}{2} \sum_a \log(\lambda_a) + \frac{1}{4} \vec{\sigma}_1^T P^{-1} \vec{\sigma}_1 - \log(f(P))$$

$$\text{where } f(P) = \int_{-\infty}^{\infty} \frac{1}{V^{(N-1)}} \exp \left( V \vec{\beta}^T \vec{1} - \frac{1}{4V} \vec{\sigma}_2^T P^{-1} \vec{\sigma}_2 \right) dV.$$

$$\begin{aligned}
\Rightarrow \frac{\partial -\log(\mathcal{Z})}{\partial P_{ij}} &= \frac{1}{2} \sum_a \frac{w_i w_j}{\lambda_a} + \frac{1}{4} \sum_{kl} \sigma_{1k} P_{ki}^{-1} P_{jl}^{-1} \sigma_{1l} - \frac{\partial \log f}{\partial P_{ij}} \\
\Rightarrow \frac{\partial -\log(\mathcal{Z})}{\partial P_{ij}} &= \frac{1}{2} \sum_a \frac{w_i w_j}{\lambda_a} + \frac{1}{4} \sum_{kl} \sigma_{1k} P^{-1} k i P_{jl}^{-1} \sigma_{1l} - \frac{1}{f} \frac{\partial f}{\partial P_{ij}}
\end{aligned}$$

$$\begin{aligned}
\frac{\partial f}{\partial P_{ij}} &= \int_{-\infty}^{\infty} \frac{1}{V^{(N-1)}} \exp \left( V \vec{\beta}^T \vec{1} - \frac{1}{4V} \vec{\sigma}_2^T P^{-1} \vec{\sigma}_2 \right) \left( -\frac{1}{4V} \sum_{kl} \sigma_{2k} P_{ki}^{-1} P_{jl}^{-1} \sigma_{2l} \right) dV \\
&= \left( -\frac{1}{4} \sum_{kl} \sigma_{2k} P_{ki}^{-1} P_{jl}^{-1} \sigma_{2l} \right) \int_{-\infty}^{\infty} \frac{1}{V^{(N)}} \exp \left( V \vec{\beta}^T \vec{1} - \frac{1}{4V} \vec{\sigma}_2^T P^{-1} \vec{\sigma}_2 \right) dV
\end{aligned}$$

$$\Rightarrow \langle z_i z_j \rangle = \frac{1}{2} \sum_a \frac{w_i w_j}{\lambda_a} + \frac{1}{4} \sum_{kl} \sigma_{1k} P^{-1} k i P_{jl}^{-1} \sigma_{1l} + \left( \frac{1}{4} \sum_{kl} \sigma_{2k} P_{ki}^{-1} P_{jl}^{-1} \sigma_{2l} \right) \frac{\int_{-\infty}^{\infty} \frac{1}{V^{(N)}} \exp \left( V \vec{\beta}^T \vec{1} - \frac{1}{4V} \vec{\sigma}_2^T P^{-1} \vec{\sigma}_2 \right) dV}{\int_{-\infty}^{\infty} \frac{1}{V^{(N-1)}} \exp \left( V \vec{\beta}^T \vec{1} - \frac{1}{4V} \vec{\sigma}_2^T P^{-1} \vec{\sigma}_2 \right) dV}$$

### 5.3 Equivalence of dynamics and Hamiltonian

**Claim 2** If  $H(\vec{v}) = -\vec{v}^T A \vec{v} - \vec{\beta}^T \vec{v}$  and  $\frac{d\vec{v}}{dt} = A \vec{v} + \vec{\beta}$  then  $\nabla H \cdot \frac{d\vec{v}}{dt} \leq 0$ .

*Proof.*

$$\begin{aligned}
H(\vec{v}) &= -\sum_{ij} A_{ij} v_i v_j - \sum_i \beta_i v_i \\
\Rightarrow \frac{\partial H}{\partial v_i} &= -\sum_{j \neq i} A_{ij} v_j - \sum_{j \neq i} A_{ji} v_j - 2A_{ii} v_i - \beta_i \\
&= -\sum_j (A_{ij} + A_{ji}^T) v_j - \beta_i \\
\Rightarrow \nabla H &= -(A + A^T) \vec{v} - \vec{\beta} \\
\Rightarrow \nabla H \cdot \frac{d\vec{v}}{dt} &= -\vec{v}^T (A + A^T) A \vec{v} - \vec{v}^T (A + A^T) \vec{\beta} - \vec{\beta}^T A \vec{v} - \vec{\beta}^T \vec{\beta} \\
&= -\vec{v}^T A^2 \vec{v} - \vec{v}^T A^T A \vec{v} - \vec{v}^T A \vec{\beta} - 2\vec{v}^T A^T \vec{\beta} - \vec{\beta}^T \vec{\beta}
\end{aligned}$$

If  $A$  is negative semi-definite, then  $A^2$  is positive semi-definite and  $A^T A$  is positive semi-definite. (Consider that  $\vec{v}^T A^T A \vec{v} = (A \vec{v})^T (A \vec{v})$  so that  $\vec{v}^T A^T A \vec{v} = 0$  if  $A \vec{v} = 0$  and  $\vec{v}^T A^T A \vec{v} < 0$  otherwise.) Therefore,  $-\vec{v}^T A^2 \vec{v} \leq 0$  and  $-\vec{v}^T A^T A \vec{v} \leq 0$ . At equilibrium,  $\vec{v}$  will be “close” to  $\vec{\beta}$  so that

$$-\vec{v}^T A \vec{\beta} - 2\vec{v}^T A^T \vec{\beta} \sim -3\vec{\beta}^T A \vec{\beta}$$

■

## 5.4 Numerically evaluating dynamics

**Claim 3** Let  $\Phi(t)$  be the fundamental matrix solution to the homogeneous differential equations  $\frac{d\vec{x}}{dt} = A\vec{x}$  so that  $\Phi(t) = V\Lambda(t)$  where the columns of  $V$  are the eigenvectors of  $A$  and

$$\Lambda(t) = \text{diag}(e^{\lambda_1 t}, \dots, e^{\lambda_n t})$$

where  $\{\lambda_i\}$  are the eigenvalues of  $A$ . (Note that if  $\lambda_i = 0$  then  $e^{\lambda_i t} = 1$  for all  $t$ .) Let  $V^\dagger$  be the pseudoinverse of  $V$ . Let  $\vec{c} = V^\dagger \vec{X}_0$ . Let

$$G(t) = \text{diag}(1/\lambda_1(1 - e^{-\lambda_1 t}), \dots, 1/\lambda_n(1 - e^{-\lambda_n t})) V^\dagger \vec{\beta}.$$

(If  $\lambda_i = 0$  then the corresponding entry of the first part of  $G(t)$  will be given by  $t$  rather than the form above.) THEN if we let  $x(t) = \Phi(t)G(t) + \Phi(t)\vec{c}$ ,  $X(t)$  solves the inhomogeneous equations  $\frac{dx}{dt} = A\vec{x} + \vec{\beta}$  with initial conditions  $x(0) = \vec{X}_0$ .

*Proof.* It is clear that

$$\frac{d\Phi\vec{c}}{dt} = A\Phi(t)\vec{c} \text{ and } \Phi(0)\vec{c} = \vec{X}_0.$$

Now,

$$\begin{aligned}
\Phi(t)G(t) &= V \begin{pmatrix} e^{\lambda_1 t} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & e^{\lambda_n t} \end{pmatrix} \begin{pmatrix} \frac{1}{\lambda_1}(1 - e^{\lambda_1 t}) & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \frac{1}{\lambda_n}(1 - e^{\lambda_n t}) \end{pmatrix} V^\dagger \vec{\beta} \\
&= V \begin{pmatrix} \frac{1}{\lambda_1}(e^{\lambda_1 t} - 1) & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \frac{1}{\lambda_n}(e^{\lambda_n t} - 1) \end{pmatrix} V^\dagger \vec{\beta} \\
\Rightarrow \frac{d\Phi(t)G(t)}{dt} &= V \begin{pmatrix} e^{\lambda_1 t} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & e^{\lambda_n t} \end{pmatrix} \\
&= V \begin{pmatrix} e^{\lambda_1 t} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & e^{\lambda_n t} \end{pmatrix} V^\dagger \vec{\beta} - V V^\dagger \vec{\beta} + \vec{\beta} \\
&= V \begin{pmatrix} e^{\lambda_1 t} - 1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & e^{\lambda_n t} - 1 \end{pmatrix} V^\dagger \vec{\beta} + \vec{\beta} \\
&= V \begin{pmatrix} \lambda_1 e^{\lambda_1 t} \cdot \frac{1}{\lambda_1}(1 - e^{-\lambda_1 t}) & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \lambda_n e^{\lambda_n t} \cdot \frac{1}{\lambda_n}(1 - e^{-\lambda_n t}) \end{pmatrix} V^\dagger \vec{\beta} + \vec{\beta} \\
&= AV\Lambda(t) \begin{pmatrix} \frac{1}{\lambda_1}(1 - e^{-\lambda_1 t}) & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \frac{1}{\lambda_n}(1 - e^{-\lambda_n t}) \end{pmatrix} V^\dagger \vec{\beta} + \vec{\beta} \\
&= AV\Lambda(t)G(t) + \vec{\beta} \\
&= A\Phi(t)G(t) + \vec{\beta}
\end{aligned}$$

Therefore  $\frac{dx}{dt} = A\Phi(t)G(t) + A\Phi(t)\vec{c} + \vec{\beta} = Ax(t) + \vec{\beta}$  and since  $G(0) = \vec{0}$ ,  $x(0) = \vec{X}_0$ . ■



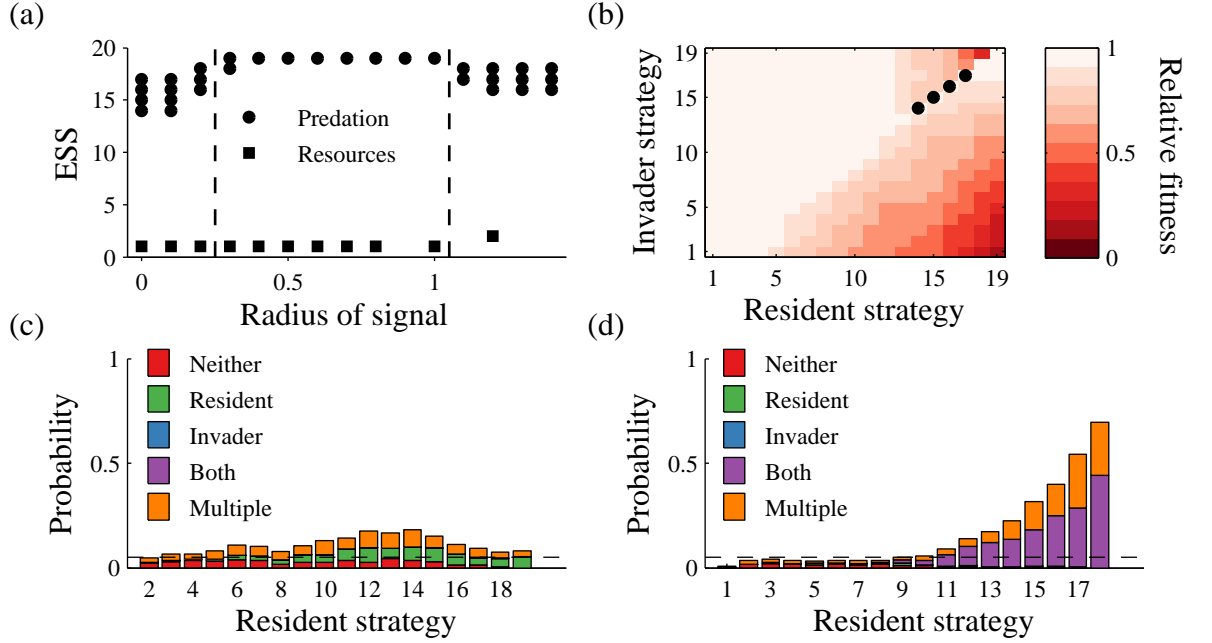


Figure 1: The ESS number of neighbors is always higher if selection is due to predation than due to resources. When selection is due to predation, there can be multiple ESS strategies, which are a non-monotonic function of the radius of the signal. In (a), we show the (possibly multiple) ESS strategies as a function of the radius of the signal. In (b) we show the relative fitness of an invader as a function of the invader and resident strategies when  $r = 0.1$ . The dots show the multiple ESSs. For (c) and (d), we choose a focal resident bird and consider the situation where neither the focal bird nor the invader has the environmental signal. The total of the stacked bars is the total probability that the invader will be predated when neither the focal bird nor the invader have the signal. We further break down these cases into five cases: there is a single informed bird and neither focal bird, only the focal resident, only the invader, or both pay attention to it or there are multiple informed birds. The horizontal dashed line shows  $1/N$ , the probability of being predated if the strategies are neutral. In (c), the invader has one fewer neighbor than the resident. In (d), the invader has one more neighbor than the resident. Parameters: in all panels  $N = 20$ ,  $\beta = 1$ ,  $T = 1$ , in (b)-(d)  $r = 0.1$ .

## 6 Figures

## 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] Roy J. Glauber. Time dependent statistics of the ising model. *Journal of Mathematical Physics*, 4(2):294–307, 1963.
- [5] Petter Holme and Gourab Ghoshal. Dynamics of networking agents competing for high centrality and low d. *Phys. Rev. Lett.*, 96(4), March 2006.

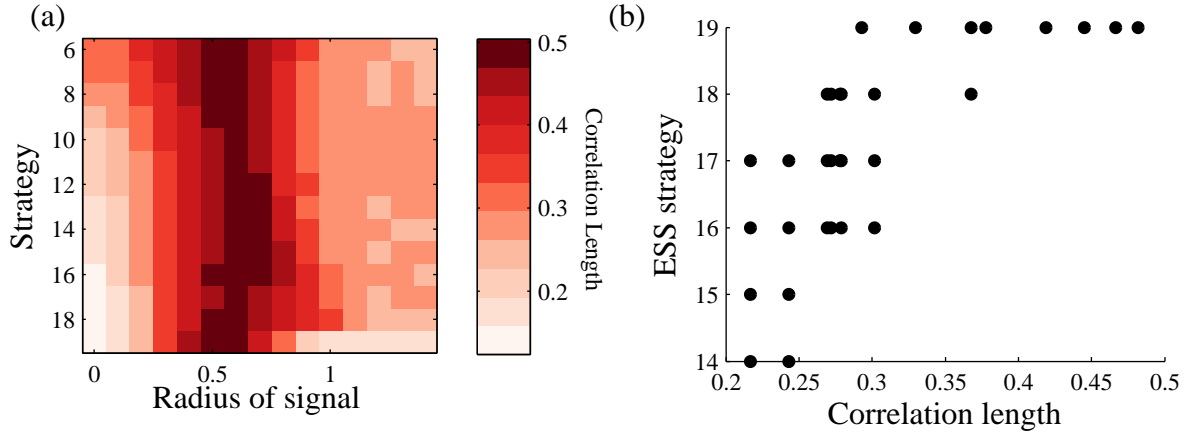


Figure 2: The correlation length of a homogeneous flock depends on both the strategy being used and the radius of the signal, but it depends much more strongly on the radius than the strategy. The ESS number of neighbors is an increasing function of correlation length of a homogeneous group. In (a), we show the average correlation length of a homogeneous flock as a function of the strategy being used and the radius of the signal. In (b), each point corresponds to a value of the radius of signal: the horizontal axis shows the correlation length of a flock paying attention to  $N/2$  neighbors for that radius and the vertical axis shows the ESS number of neighbors for that radius. Parameters:  $N = 20$ ,  $\beta = 1$ ,  $T = 1$ .

- [6] Carl Meyer. *Matrix Analysis and Applied Linear Algebra*. SIAM: Society for Industrial and Applied Mathematics, 2001.
- [7] Martin Nowak. *Evolutionary Dynamics*. Belknap / Harvard, 2006.
- [8] 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.
- [9] George Forrest Young, Luca Scardovi, and Naomi Ehrich Leonard. Robustness of noisy consensus dynamics with directed communication. *Proceedings of the American Control Conference*, 2010.

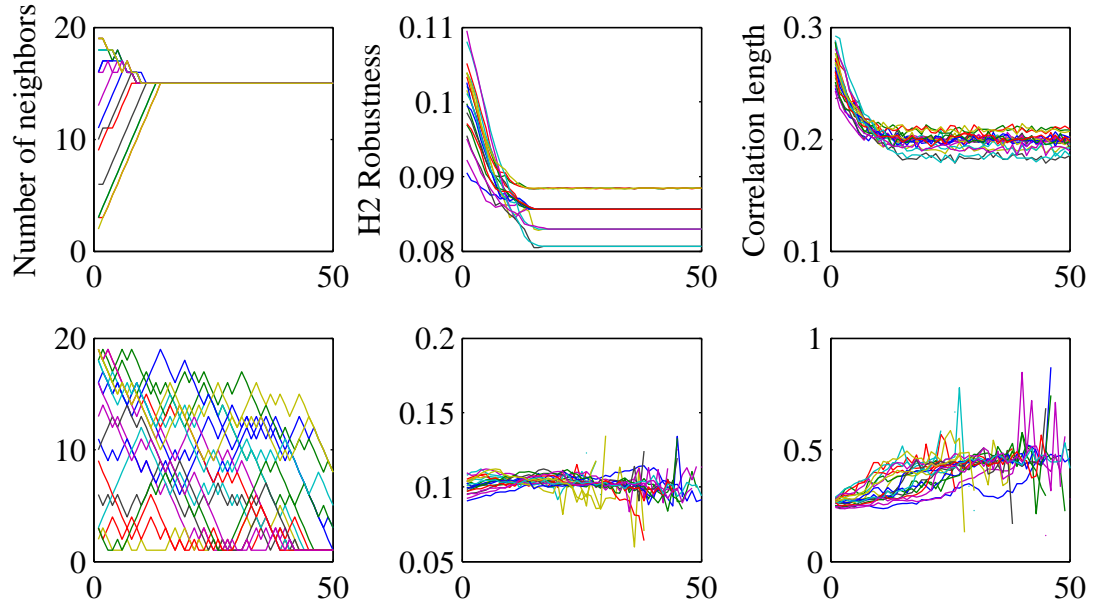


Figure 3: When the birds choose the optimal number of neighbors, given the strategies the rest of the flock are using, they settle on higher strategies when selection is due to predation than when selection is due to resources. Under selection due to predation, the  $\mathcal{H}_2$  robustness and correlation length of the flock tend to decrease, whereas under selection due to resources,  $\mathcal{H}_2$  robustness is relatively constant and correlation length tends to increase. The upper row shows results from implementing selection based to predation and the lower row shows results from implementing selection based on 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 many initial conditions, how the  $\mathcal{H}_2$  robustness changes over time. The third column shows, for many initial conditions, how the correlation length changes over time. Parameters:  $N = 20$ ,  $\beta = 1$ ,  $r = .1$ ,  $T = 1$ .