Supporting Information

Bialek et al. 10.1073/pnas.1324045111

SI Text

I. Data

The data that we analyze here were obtained from observations on large flocks of starlings, *Sturnus vulgaris*, in the field. Using stereometric photography and innovative computer vision techniques (1, 2) the individual 3D coordinates and velocities were measured in cohesive groups of up to 4,268 individuals (3, 4). As summarized in Table S1, we have data from 21 distinct flocking events, with flock sizes ranging from 122 to 4,268 individuals and linear extensions from 9.1 to 85.7 m. Each event consists of up to 80 consecutive 3D configurations (individual positions and velocities) at time intervals of $1/10$ s. All events correspond to strongly ordered flocks, with polarization [from Eq. **13** of the main text] between $|\vec{\mathbf{P}}| = 0.844$ and $|\vec{\mathbf{P}}| = 0.992$. The border of each flock at each instant of time has been computed using the α-shape algorithm (5), as explained in detail in ref. 2.

II. The Maximum Entropy Approach

The concept of entropy has its roots in thermodynamics, roughly 150 y ago. The idea that we can use maximum entropy as a strategy to construct simplified models outside of equilibrium thermodynamics is now more than 50 y old (6). Here, so that our discussion is self-contained, we review this general strategy (see also ref. 7 and appendix A.7 in ref. 8).

We assume that the state of the system can be described by a set of variables that we shall call $\mathbf{v} \equiv \{\vec{\mathbf{v}}_1, \vec{\mathbf{v}}_2 \cdots, \vec{\mathbf{v}}_N\}$, by analogy with the velocities of birds in a flock. Although we can measure, for example, the velocity of every bird in a flock, we typically can not collect enough data to make reliable estimates of very complicated quantitates. As an example, with N variables describing the state of the system, we need more than N independent measurements to be sure that the covariance matrix of these variables is not artificially singular. What does seem reasonable is to assume that there is a much smaller set of observables, $\{O_\mu(\mathbf{v})\}$ with $\mu = 1, 2, \cdots, K$, that we can extract from the system, and that we have enough data to make reliable statements about the average values of these observables, $\{\langle O_\mu(\mathbf{v})\rangle_{\rm exp}\}$.

Our task is to build a probability distribution $P(\mathbf{v})$ such that we reproduce, exactly, the expectation values of the K observables, that is

$$\langle O_\mu(\mathbf{v})\rangle_P \equiv \sum_{\mathbf{v}} P(\mathbf{v}) O_\mu(\mathbf{v}) = \langle O_\mu(\mathbf{v})\rangle_{\rm exp}, \qquad \text{[S1]}$$

for all $\mu = 1, 2, \cdots, K$; it is useful to phrase the normalization of the distribution as a similar constraint, the statement that the average of the function $O_0(\mathbf{v}) = 1$ must equal the experimental value of 1.

The problem is that that there are infinitely many distributions that can satisfy the constraints in Eq. **S1**. Out of all these distributions, we want to find the one that has as little structure as possible, so that we can derive the minimal consequences of the experimental observations $\{\langle O_\mu(\mathbf{v})\rangle_{\rm exp}\}$. Asking for a probability distribution $P(\mathbf{v})$ that has as little structure as possible is equivalent to asking that the variables $\mathbf{v}$ that we draw out of this distribution be as random as possible. Shannon proved that the only measure of (lack of) structure or randomness that is consistent with several simple constraints is the entropy of the distribution (9, 10),

$$S[P] = -\sum_{\mathbf{v}} P(\mathbf{v}) \ln P(\mathbf{v}). \qquad \text{[S2]}$$

Thus, we are looking for the distribution $P(\mathbf{v})$ that maximizes the entropy in Eq. **S2** while obeying the experimental constraints from Eq. **S1**. Such constrained optimization problems can be solved using the method of the Lagrange multipliers (11): we introduce a generalized entropy function,

$$\mathcal{S}[P; \{\lambda_\nu\}] = S[P] - \sum_{\mu=0}^{K} \lambda_\mu \left[\langle O_\mu(\mathbf{v})\rangle_P - \langle O_\mu(\mathbf{v})\rangle_{\rm exp}\right], \qquad \text{[S3]}$$

where a multiplier λ_μ appears for each constraint to be satisfied, and then we maximize $\mathcal{S}$ both with respect to the probability distribution $P(\mathbf{v})$ and with respect to the parameters $\{\lambda_\mu\}$.

Maximizing with $\mathcal{S}$ respect to $P(\mathbf{v})$ gives

$$P(\mathbf{v}) = \frac{1}{\mathcal{Z}(\{\lambda_\nu\})} \exp\left[-\sum_{\mu=1}^{K} \lambda_\mu O_\mu(\mathbf{v})\right], \qquad \text{[S4]}$$

where $\mathcal{Z}(\{\lambda_\nu\}) = \exp(\lambda_0 - 1)$. Since optimizing with respect to λ_0 will enforce normalization of the distribution, we can write, explicitly,

$$\mathcal{Z}(\{\lambda_\nu\}) = \sum_{\mathbf{v}} \exp\left[-\sum_{\mu=1}^{K} \lambda_\mu O_\mu(\mathbf{v})\right]. \qquad \text{[S5]}$$

Maximizing with respect to $\{\lambda_\nu\}$ gives us the set of K simultaneous equations in Eq. **S1**, which we can now write more explicitly as

$$\langle O_\mu(\mathbf{v})\rangle_{\rm exp} = \frac{1}{\mathcal{Z}(\{\lambda_\nu\})} \sum_{\mathbf{v}} O_\mu(\mathbf{v}) \exp\left[-\sum_{\nu=1}^{K} \lambda_\nu O_\nu(\mathbf{v})\right]. \qquad \text{[S6]}$$

We note that, in general, this is a very nonlinear set of equations for the parameters $\{\lambda_\nu\}$, and very hard to solve. In section IV we exploit special features of the flock problem—in particular, the strong polarization of the flock—to simplify this problem so that we can make analytic progress.

Maximum entropy distributions are mathematically equivalent to the Boltzmann distribution in statistical physics. We recall that if a physical system in state $\mathbf{v}$ has energy $E(\mathbf{v})$, then when it comes to equilibrium at temperature T the probability that is in any particular state is given by

$$P_{\rm Boltz}(\mathbf{v}) = \frac{1}{Z} \exp\left[-\frac{E(\mathbf{v})}{k_B T}\right], \qquad \text{[S7]}$$

where k_B is Boltzmann's constant, and serves to convert between conventional units of temperature and energy. Comparing with Eq. **S4**, we see that the maximum entropy distribution is equivalent to a Boltzmann distribution with an effective energy

$$\frac{E(\mathbf{v})}{k_B T} = \sum_{\mu=1}^{K} \lambda_\mu O_\mu(\mathbf{v}). \qquad \text{[S8]}$$

We note that this energy is the sum of several terms, one for each of the observables whose expectation value we fix based on experimental data.

It also is useful to note the connection of the maximum entropy approach with more conventional model building. If we take the form of the probability distribution in Eq. **S4** as given, then our problem is only to fit the parameters $\{\lambda_\nu\}$. A standard method is maximum likelihood. If we have N_s independent samples of the system's state, $\mathbf{v}^{(1)}, \mathbf{v}^{(2)}, \cdots, \mathbf{v}^{(N_s)}$, then the probability that the model generates these data is given by

$$P_{\rm model}({\rm data}) = \prod_{i=1}^{N_s} P\left(\mathbf{v}^{(i)}\right). \quad \textbf{[S9]}$$

Substituting from Eq. **S4** we can make this more explicit:

$$\begin{aligned} P_{\rm model}({\rm data}) &= \frac{1}{\mathcal{Z}^{N_s}(\{\lambda_\nu\})} \prod_{i=1}^{N_s} \exp\left[-\sum_{\mu=1}^{K} \lambda_\mu O_\mu\left(\mathbf{v}^{(i)}\right)\right] \\ &= \frac{1}{\mathcal{Z}^{N_s}(\{\lambda_\nu\})} \exp\left[-\sum_{\mu=1}^{K} \lambda_\mu \sum_{i=1}^{N_s} O_\mu\left(\mathbf{v}^{(i)}\right)\right]. \end{aligned} \quad \textbf{[S10]}$$

Then we can form the normalized log probability,

$$\frac{1}{N_s} \ln P_{\rm model}({\rm data}) = -\ln \mathcal{Z}(\{\lambda_\nu\}) - \sum_{\mu=1}^{K} \lambda_\mu \left[\frac{1}{N_s} \sum_{i=1}^{N_s} O_\mu\left(\mathbf{v}^{(i)}\right)\right] \quad \textbf{[S11]}$$

$$= -\ln \mathcal{Z}(\{\lambda_\nu\}) - \sum_{\mu=1}^{K} \lambda_\mu \langle O_\mu(\mathbf{v})\rangle_{\rm exp}, \quad \textbf{[S12]}$$

where in the last step we recognize the normalized sum over samples as the experimental expectation value. Now if we want to maximize the probability, or likelihood, we should differentiate with respect to the parameters and set the result to zero:

$$\frac{\partial \ln P_{\rm model}({\rm data})}{\partial \lambda_\mu} = 0 \Rightarrow \frac{\partial \ln \mathcal{Z}(\{\lambda_\nu\})}{\partial \lambda_\mu} = -\langle O_\mu(\mathbf{v})\rangle_{\rm exp}. \quad \textbf{[S13]}$$

However, with the explicit expression for $\mathcal{Z}$ in Eq. **S5**, we can compute

$$\begin{aligned} \frac{\partial \ln \mathcal{Z}(\{\lambda_\nu\})}{\partial \lambda_\mu} &= \frac{1}{\mathcal{Z}(\{\lambda_\nu\})} \frac{\partial \mathcal{Z}(\{\lambda_\nu\})}{\partial \lambda_\mu} \\ &= \frac{1}{\mathcal{Z}(\{\lambda_\nu\})} \frac{\partial}{\partial \lambda_\mu} \sum_{\mathbf{v}} \exp\left[-\sum_{\nu=1}^{K} \lambda_\nu O_\nu(\mathbf{v})\right] \end{aligned} \quad \textbf{[S14]}$$

$$= -\frac{1}{\mathcal{Z}(\{\lambda_\nu\})} \sum_{\mathbf{v}} \exp\left[-\sum_{\nu=1}^{K} \lambda_\nu O_\nu(\mathbf{v})\right] O_\mu(\mathbf{v}) \quad \textbf{[S15]}$$

$$= -\sum_{\mathbf{v}} \frac{1}{\mathcal{Z}(\{\lambda_\nu\})} \exp\left[-\sum_{\nu=1}^{K} \lambda_\nu O_\nu(\mathbf{v})\right] O_\mu(\mathbf{v}) \quad \textbf{[S16]}$$

$$= -\sum_{\mathbf{v}} P(\mathbf{v}) O_\mu(\mathbf{v}). \quad \textbf{[S17]}$$

We recognize this as the expectation value of $O_\mu(\mathbf{v})$ with respect to the probability distribution $P(\mathbf{v})$. Thus, we have

$$\frac{\partial \ln \mathcal{Z}(\{\lambda_\nu\})}{\partial \lambda_\mu} = -\langle O_\mu(\mathbf{v})\rangle_P, \quad \textbf{[S18]}$$

and hence Eq. **S13** becomes

$$\langle O_\mu(\mathbf{v})\rangle_P = \langle O_\mu(\mathbf{v})\rangle_{\rm exp}. \quad \textbf{[S19]}$$

That is, once we have the form of the maximum entropy distribution in Eq. **3** of the main text, maximizing the likelihood of the data with respect to parameters is equivalent to imposing the constraints in Eq. **S1**.

III. Maximum Entropy Model for Flocks

Let us now apply the maximum entropy approach to the case of bird flocks. The state of the system is characterized by the set $\mathbf{v} \equiv \{\vec{\mathbf{v}}_1, \vec{\mathbf{v}}_2 \cdots, \vec{\mathbf{v}}_N\}$ of the individual bird velocities. As discussed in the main text, we consider observables that measure the local correlations between birds and their neighbors, and the mean and variance of flight speeds.

When we look at a snapshot of the flock, we can identify bird j as being in the neighborhood of bird i ($j \in \mathcal{N}_i$) if it is one of the closest n_c neighbors. Then we measure the mean square difference in velocity between a bird and those in its neighborhood,

$$Q_{\rm int} = \frac{1}{2Nv_0^2} \sum_{i=1}^{N} \frac{1}{n_c} \sum_{j\in\mathcal{N}_i} |\vec{\mathbf{v}}_i - \vec{\mathbf{v}}_j|^2, \quad \textbf{[S20]}$$

where we have normalized by a scale v_0 to obtain a dimensionless measure; in solving the model we shall see that it is natural to set this scale equal to the observed mean speed of the birds. It will be convenient to write this in a slightly different form, so we introduce matrix $\hat{n}_{ij} = 1$ if $j \in \mathcal{N}_i$ and $\hat{n}_{ij} = 0$ otherwise. Then we have

$$Q_{\rm int} = \frac{1}{2Nv_0^2} \frac{1}{n_c} \sum_{i=1}^{N} \sum_{j=1}^{N} \hat{n}_{ij} |\vec{\mathbf{v}}_i - \vec{\mathbf{v}}_j|^2. \quad \textbf{[S21]}$$

We notice that the indices i and j appear symmetrically, but the matrix $\hat{n}_{ij}$ is not symmetric, since being in the neighborhood is not a symmetrical relationship (if you are my nearest neighbor, I might not be your nearest neighbor). Only the symmetric part survives the summation, so we can write

$$Q_{\rm int} = \frac{1}{2Nv_0^2} \frac{1}{n_c} \sum_{i=1}^{N} \sum_{j=1}^{N} n_{ij} |\vec{\mathbf{v}}_i - \vec{\mathbf{v}}_j|^2, \quad \textbf{[S22]}$$

where $n_{ij} = (\hat{n}_{ij} + \hat{n}_{ji})/2$.

In addition to $Q_{\rm int}$, we chose as observables the mean speed and the mean square speed across the flock,

$$V = \frac{1}{N} \sum_{i=1}^{N} v_i \quad \textbf{[S23]}$$

$$V_2 = \frac{1}{N} \sum_{i=1}^{N} v_i^2, \quad \textbf{[S24]}$$

where $v_i = |\vec{\mathbf{v}}_i|$ is the speed of bird i.

Eq. **S8** tells us that the effective energy function or Hamiltonian for a maximum entropy model is composed of one term for each of the observables whose expectation values we match to the data. Thus, we should have

$$\mathcal{H}(\mathbf{v}) = \lambda_1 Q_{\rm int} + \lambda_2 V + \lambda_3 V_2, \quad \textbf{[S25]}$$

and the probability distribution

$$P(\mathbf{v}) = \frac{e^{-\mathcal{H}(\mathbf{v})}}{\mathcal{Z}(\lambda_1, \lambda_2, \lambda_3)}. \quad \textbf{[S26]}$$

It will be useful to absorb factors of N so that the effective energy becomes extensive, that is proportional (on average) to the number of birds in the flock, while the parameters of the model remain formally independent of N. Similarly, we would like to separate the choice of units for velocity from the dimensionless parameters of our model, so we introduce a scale v_0 as above. Thus, we write

$$\mathcal{H}(\mathbf{v}) = \frac{J}{4v_0^2}\sum_{i,j=1}^{N} n_{ij}|\vec{\mathbf{v}}_i - \vec{\mathbf{v}}_j|^2 + \frac{g}{2v_0^2}\sum_{i=1}^{N} v_i^2 - \frac{\mu}{v_0}\sum_{i=1}^{N} v_i. \qquad \textbf{[S27]}$$

With $P(\mathbf{v}) \propto \exp[-\mathcal{H}(\mathbf{v})]$, we obtain Eq. **3** of the main text.

IV. Solving the Model

Maximum entropy methods involve building models grounded in measurements of average quantities. Since the state of our system is defined by the velocities of all of the individual birds, it might seem that to define an average we need many snapshots of the entire flock. However, the quantities that interest us are local—they describe the behavior of individual birds (V_1 and V_2) or of individuals relative to their neighbors ($Q_{\rm int}$). Thus, we can compute averages of these local operators over all of the birds in the flock, and since the flocks are large we expect that this will be equivalent to averaging over an ensemble of snapshots. Indeed, we shall see that the expectation values computed over the many birds in a single snapshot do not fluctuate much from moment to moment, and correspondingly the parameters J and g also do not vary systematically with time during the flocking events, or even across different flocks. This is similar to what we found in the analysis of directional ordering alone (7).

The first step in using the maximum entropy model is to compute the partition function $\mathcal{Z}(J,g,\mu)$. Since the role of $\mathcal{Z}(J,g,\mu)$ is to enforce normalization, we have

$$\mathcal{Z}(J,g,\mu) = \int d\mathbf{v}\, e^{-\mathcal{H}(\mathbf{v})}, \qquad \textbf{[S28]}$$

where $d\mathbf{v}$ is the volume element in the space of all of the (3D) velocities, $d\mathbf{v} = \prod_i d^3\vec{\mathbf{v}}_i$.

A. Computation with Free Boundary Conditions. We begin by treating all birds as equivalent, without regard to their location in the interior or on the boundary of the flock, and we return to this below. It will be useful to think of the velocity as being composed of a speed and a direction, $\vec{\mathbf{v}}_i = v_i\vec{\mathbf{s}}_i$, where $|\vec{\mathbf{s}}_i| = 1$. Translating into these variables, we obtain from Eq. **S27**:

$$\mathcal{H}(\mathbf{v}) = \frac{J}{4v_0^2}\sum_{i,j=1}^{N} n_{ij}|v_i\vec{\mathbf{s}}_i - v_j\vec{\mathbf{s}}_j|^2 + \frac{g}{2v_0^2}\sum_{i=1}^{N} v_i^2 - \frac{\mu}{v_0}\sum_{i=1}^{N} v_i \qquad \textbf{[S29]}$$

$$= \frac{J}{4v_0^2}\sum_{i,j=1}^{N} n_{ij}\left[v_i^2 - 2v_iv_j\vec{\mathbf{s}}_i\cdot\vec{\mathbf{s}}_j + v_j^2\right] + \frac{g}{2v_0^2}\sum_{i=1}^{N} v_i^2 - \frac{\mu}{v_0}\sum_{i=1}^{N} v_i \qquad \textbf{[S30]}$$

$$= -\frac{J}{2v_0^2}\sum_{i,j=1}^{N} n_{ij}v_iv_j\vec{\mathbf{s}}_i\cdot\vec{\mathbf{s}}_j + \frac{1}{2v_0^2}\sum_{i=1}^{N}\left(g + J\sum_{k=1}^{N} n_{ik}\right)v_i^2 - \frac{\mu}{v_0}\sum_{i=1}^{N} v_i. \qquad \textbf{[S31]}$$

Notice that the term controlling the mean square speed now has two contributions, one from the direct control parameter g and one from the social interactions with neighbors, $\propto J$.

In addition to rewriting the Hamiltonian, we also need to express the volume element $d\mathbf{v}$ in terms of the new direction and speed variables. For each bird,

$$d^3\mathbf{v}_i = v_i^2 dv_i d^3\vec{\mathbf{s}}_i\delta(|\vec{\mathbf{s}}_i| - 1), \qquad \textbf{[S32]}$$

where the delta function enforces the constraint that $\vec{\mathbf{s}}_i$ is a unit vector, and the factor v_i^2 is the Jacobian of the transformation. In the limit that speed fluctuations are small—which they are in the flock—the effect of the Jacobian can always be absorbed into a redefinition of the parameters μ and g, so we drop this term here. Thus, we have

$$\mathcal{Z}(J,g,\mu) = \int\prod_{i=1}^{N} dv_i d^3\vec{\mathbf{s}}_i\delta(|\vec{\mathbf{s}}_i| - 1)\exp\left[\frac{J}{2v_0^2}\sum_{i,j=1}^{N} n_{ij}v_iv_j\vec{\mathbf{s}}_i\cdot\vec{\mathbf{s}}_j - \frac{1}{2v_0^2}\sum_{i=1}^{N}\left(g + J\sum_{k=1}^{N} n_{ik}\right)v_i^2 + \frac{\mu}{v_0}\sum_{i=1}^{N} v_i\right]. \qquad \textbf{[S33]}$$

Now we want to use the fact that fluctuations are small to simplify our calculation; we can verify, at the end, that the fluctuations predicted by the model really are small, and hence that our approximations are consistent. This is a now classical approximation scheme in the theory of magnetism (12), but we go through the details here in the hopes of making the calculation accessible to a broader audience.

We can write the speeds as

$$v_i = V(1 + \epsilon_i), \qquad \textbf{[S34]}$$

where V is the mean speed over the flock from Eq. **S23**,

$$V = \frac{1}{N}\sum_{i=1}^{N} v_i, \qquad \textbf{[S35]}$$

and ϵ_i is the fractional fluctuation around this mean; we expect $|\epsilon_i| \ll 1$. Notice that with this definition we have

$$\sum_{i=1}^{N}\epsilon_i = 0. \qquad \textbf{[S36]}$$

Transforming from integrating over speeds to integrating over their fluctuations, we have

$$\prod_{i=1}^{N} dv_i = V^N dV\left(\prod_{i=1}^{N} d\epsilon_i\right)\delta\left(\sum_{j=1}^{N}\epsilon_j\right). \qquad \textbf{[S37]}$$

To say that fluctuations in direction are small requires a bit more care. We can average the unit vectors $\vec{\mathbf{s}}_i$ to obtain the polarization of the flock as in Eq. **13** of the main text,

$$\vec{\mathbf{P}} = \frac{1}{N}\sum_{i=1}^{N}\vec{\mathbf{s}}_i. \qquad \textbf{[S38]}$$

This polarization has a magnitude P and a direction that we will denote by the unit vector $\hat{\mathbf{n}}$, so that $\vec{\mathbf{P}} = P\hat{\mathbf{n}}$. We expect that flight directions of individual birds will be close to $\hat{\mathbf{n}}$, so we can write

$$\vec{\mathbf{s}}_i = s_i^L\hat{\mathbf{n}} + \vec{\pi}_i, \qquad \textbf{[S39]}$$

where $\vec{\pi}_i$ is a (small) vector perpendicular to $\hat{\mathbf{n}}$, and the longitudinal term s_i^L is necessary to be sure that $\vec{\mathbf{s}}_i$ remains a unit vector. As with the ϵ_i above, not all N of these variables are independent, since the definition of the polarization in Eq. **S38** requires that

$$P = \frac{1}{N}\sum_{i=1}^{N} s_i^L, \tag{S40}$$

and

$$\sum_{i=1}^{N} \vec{\pi}_i = 0. \tag{S41}$$

Thus, we have

$$\begin{aligned}\prod_{i=1}^{N} d^3\vec{s}_i\,\delta(|\vec{s}_i|-1) &= \int \frac{d^2\vec{n}}{4\pi}\int dP\left[\prod_{i=1}^{N} d^2\vec{\pi}_i ds_i^L\,\delta\left(\sqrt{[s_i^L]^2+|\vec{\pi}_i|^2}-1\right)\right]\\ &\quad \delta\left(P-\frac{1}{N}\sum_{i=1}^{N} s_i^L\right)\delta\left(\sum_{i=1}^{N}\vec{\pi}_i\right).\end{aligned} \tag{S42}$$

Now, if we substitute into Eq. **S31**, we have

$$\begin{aligned}\mathcal{H}(\mathbf{v}) = &-\frac{JV^2}{2v_0^2}\sum_{i,j=1}^{N} n_{ij}(1+\epsilon_i)(1+\epsilon_j)\left(s_i^L s_j^L + \vec{\pi}_i\cdot\vec{\pi}_j\right)\\ &+\frac{V^2}{2v_0^2}\sum_{i=1}^{N}\left(g+J\sum_{k=1}^{N} n_{ik}\right)(1+\epsilon_i)^2 - \frac{N\mu}{v_0}V.\end{aligned} \tag{S43}$$

Although we have changed variables in a way that makes it easy to make the approximation that fluctuations are small, we have not actually used this approximation yet in simplifying the Hamiltonian.

We notice that one set of delta functions in Eq. **S42** enforces

$$s_i^L = \sqrt{1-|\vec{\pi}_i|^2} \approx 1 - |\vec{\pi}_i|^2/2 + \cdots, \tag{S44}$$

where we now use the approximation that $|\vec{\pi}_i|$ is small. If we substitute this into Eq. **S43**, then to be consistent we should keep only terms up to second order in $\vec{\pi}_i$ and ϵ_i. The result is

$$\begin{aligned}\mathcal{H}(\mathbf{v}) = &-\frac{JV^2}{2v_0^2}\sum_{i,j=1}^{N} n_{ij}(1+\epsilon_i)(1+\epsilon_j)\\ &+\frac{V^2}{2v_0^2}\sum_{i=1}^{N}\left(g+J\sum_{k=1}^{N} n_{ik}\right)(1+\epsilon_i)^2 - \frac{N\mu}{v_0}V\\ &-\frac{JV^2}{2v_0^2}\sum_{i,j=1}^{N} n_{ij}\left(-|\vec{\pi}_i|^2/2 - |\vec{\pi}_j|^2/2 + \vec{\pi}_i\cdot\vec{\pi}_j\right).\end{aligned} \tag{S45}$$

A crucial simplification is that the terms related to speed fluctuations (ϵ_i) are decoupled from those related to directional fluctuations ($\vec{\pi}_i$). Thus, we have, as in Eq. **6** of the main text,

$$\mathcal{H}(\mathbf{v}) = \mathcal{H}_{dir}(\{\vec{\pi}_i\}) + \mathcal{H}_{sp}(\{\epsilon_i\}) + E_0(V), \tag{S46}$$

where $E_0(V)$ is the effective energy when all $\epsilon_i = 0$,

$$E_0(V) = N\left(\frac{gV^2}{2v_0^2} - \mu V\right). \tag{S47}$$

Collecting terms, and dropping constants independent of $\{\vec{\pi}_i\}$ and $\{\epsilon_i\}$, we find that

$$\mathcal{H}_{dir}(\{\vec{\pi}_i\}) = \frac{JV^2}{2v_0^2}\sum_{i,j=1}^{N} N_{ij}\vec{\pi}_i\cdot\vec{\pi}_j \tag{S48}$$

$$\mathcal{H}_{sp}(\{\epsilon_i\}) = \frac{V^2}{2v_0^2}\sum_{i,j=1}^{N}\left(g\delta_{ij} + JN_{ij}\right)\epsilon_i\epsilon_j, \tag{S49}$$

where the matrix N_{ij} has the form

$$N_{ij} = -n_{ij} + \delta_{ij}\sum_{k=1}^{N} n_{ik}. \tag{S50}$$

In trying to compute the partition function, we will need to integrate not just over the local variables $\{\epsilon_i, \vec{\pi}_i\}$, but also—as can be seen from the volume elements in Eqs. **S37** and **S42**—over the global variables V, P, and $\hat{\mathbf{n}}$. The integral over the direction of polarization is simple because there is no dependence of the integrand on $\hat{\mathbf{n}}$; this is a consequence of the overall rotational invariance in our formulation of the problem. The integral over the magnitude of the polarization is also simple, since the delta function just gives us

$$P = \frac{1}{N}\sum_{i=1}^{N} s_i^L \approx 1 - \frac{1}{2N}\sum_{i=1}^{N}|\vec{\pi}_i|^2. \tag{S51}$$

The integral over V is more interesting, since the V dependence of the integrand is dominated by $E_0(V)$. Thus, we need to do an integral of the form

$$\mathcal{Z}_V \approx \int dV\, e^{-E_0(V)}. \tag{S52}$$

The key point is that $E_0 \propto N$, and so the integrand is very sharply peaked around some V_*. However, the average of V is one of the quantities that we are fixing from the data, so we must have $V_* = \langle V\rangle_{exp}$, and this serves to set the parameter μ, as explained in the main text. Importantly, the factor of N ensures that the variations in V around V_* will be very small in large flocks, and hence we can replace $V \to V_* = \langle V\rangle_{exp}$ everywhere else in our calculations. We are also free to choose the scale $v_0 = \langle V\rangle_{exp}$, and then we can simplify

$$\mathcal{H}_{dir}(\{\vec{\pi}_i\}) = \frac{J}{2}\sum_{i,j=1}^{N} N_{ij}\vec{\pi}_i\cdot\vec{\pi}_j, \tag{S53}$$

$$\mathcal{H}_{sp}(\{\epsilon_i\}) = \frac{1}{2}\sum_{i,j=1}^{N}\left(g\delta_{ij} + JN_{ij}\right)\epsilon_i\epsilon_j. \tag{S54}$$

This separation of direction and speed variables in the Hamiltonian means that the partition function can be factorized,

$$\mathcal{Z}(J,g,\mu) \propto \mathcal{Z}_{dir}(J)\,\mathcal{Z}_{sp}(J,g)e^{Ng/2}, \tag{S55}$$

where

$$\mathcal{Z}_{dir}(J) = \int\left[\prod_{i=1}^{N} d^2\vec{\pi}_i\right]\delta\left(\sum_{i=1}^{N}\vec{\pi}_i\right)e^{-\mathcal{H}_{dir}(\{\vec{\pi}_i\})} \tag{S56}$$

$$\mathcal{Z}_{\rm sp}(J,g) = \int \left[\prod_{\rm i=1}^{N} d\epsilon_{\rm i}\right] \delta\left(\sum_{\rm j=1}^{N} \epsilon_{\rm j}\right) e^{-\mathcal{H}_{\rm sp}(\{\epsilon_{\rm i}\})}. \qquad \textbf{[S57]}$$

Now we have to do the integrals in Eqs. **S56** and **S57**, but these are not very difficult because they are Gaussians. The behavior of these integrals is determined by the structure of the matrix $N_{\rm ij}$. To understand this structure, imagine that the birds are in a line, and the relevant neighborhood is just the two nearest neighbors along the line. Then we can see that $N_{\rm ij}$ is the discrete approximation to the (negative) second derivative along the line. In higher dimensions this becomes the Laplacian operator, and so $N_{\rm ij}$ is called a "Laplacian matrix." As with the negative Laplacian, the eigenvalues $\{\Lambda_{\rm a}\}$ of $N_{\rm ij}$ are positive, except for the smallest one, which is exactly zero ($\Lambda_1 = 0$). If we define the eigenvectors of $N_{\rm ij}$ by $w_{\rm i}^{\rm a}$ such that

$$\sum_{\rm j=1}^{N} N_{\rm ij} w_{\rm j}^{\rm a} = \Lambda_{\rm a} w_{\rm i}^{\rm a}, \qquad \textbf{[S58]}$$

then the eigenvector associated with the zero eigenvalue is the uniform mode, $w_{\rm i}^1 = $ constant. However, displacements along this direction are fixed to zero by the delta functions that appear in the integrals of Eqs. **S56** and **S57**, and this is crucial for doing the integrals.

We recall that, for a general $N \times N$ matrix $M_{\rm ij}$,

$$\int d^N x \, \exp\left(-\frac{1}{2}\sum_{\rm i,j=1}^{N} x_{\rm i} M_{\rm ij} x_{\rm j}\right) = \left[\frac{(2\pi)^N}{\det M}\right]^{1/2} \propto \exp\left(-\frac{1}{2}\sum_{\rm a=1}^{N} \ln[\lambda_{\rm a}(M)]\right), \qquad \textbf{[S59]}$$

where $\lambda_{\rm n}(M)$ are the eigenvalues of M. In the case of $\mathcal{Z}_{\rm sp}$, we have

$$\mathcal{Z}_{\rm sp}(J,g) = \int \left[\prod_{\rm i=1}^{N} d\epsilon_{\rm i}\right] \delta\left(\sum_{\rm j=1}^{N} \epsilon_{\rm j}\right) \exp\left[-\frac{1}{2}\sum_{\rm i,j=1}^{N} \epsilon_{\rm i}\left(g\delta_{\rm ij} + JN_{\rm ij}\right)\epsilon_{\rm j}\right]. \qquad \textbf{[S60]}$$

The relevant matrix is now $M_{\rm ij} = g\delta_{\rm ij} + JN_{\rm ij}$, and the eigenvalues are $\lambda_{\rm a}(M) = g + J\Lambda_{\rm a}$, where again $\Lambda_{\rm a}$ are the eigenvalues of the Laplacian matrix $N_{\rm ij}$. We note that the integral runs over N dimensions, but the delta function fixes one combination of the $\{\epsilon_{\rm i}\}$ to be zero, and as noted above this combination is parallel to the first eigenvector. So, up to constant factors, the effect of the delta function is to exclude the first (zero) eigenvalue from the sum in Eq. **S59**, so that

$$\mathcal{Z}_{\rm sp}(J,g) \propto \exp\left(-\frac{1}{2}\sum_{\rm a=2}^{N} \ln[g + J\Lambda_{\rm a}]\right). \qquad \textbf{[S61]}$$

Since the effective Hamiltonian for speed fluctuations in Eq. **S54** is a quadratic function of the $\{\epsilon_{\rm i}\}$, the probability distribution of the speed fluctuations is Gaussian,

$$P(\{\epsilon_{\rm i}\}) = \frac{1}{\mathcal{Z}_{\rm sp}(J,g)} \delta\left(\sum_{\rm j=1}^{N} \epsilon_{\rm j}\right) \exp\left[-\frac{1}{2}\sum_{\rm i,j=1}^{N} \epsilon_{\rm i} M_{\rm ij} \epsilon_{\rm j}\right]. \qquad \textbf{[S62]}$$

Thus, we can calculate the correlations between the values of ϵ for different birds i and j in a standard way: we rotate our coordinates into the eigenvectors of the matrix $M_{\rm ij}$, and note that in this basis fluctuations along each coordinate are independent with variance $1/\lambda_{\rm n}(M)$; to recover the correlations in the original basis we rotate back. Again we have to be careful to respect the delta function, which serves to eliminate the fluctuations along $w_{\rm i}^1$. The end result is that

$$\langle \epsilon_{\rm i} \epsilon_{\rm j} \rangle = \sum_{\rm a=2}^{N} \frac{w_{\rm i}^{\rm a} w_{\rm j}^{\rm a}}{g + J\Lambda_{\rm a}}. \qquad \textbf{[S63]}$$

This result, or more precisely its generalization to the case where we treat the birds on the boundary of the flock separately, Eq. **S90**, is the basis for our prediction of the speed correlations as a function of the distance between birds, in Fig. 2*C* of the main text.

We can carry through the same calculation for the direction fluctuations. The only differences are that the vector $\vec{\pi}_{\rm i}$ has two components, so there are twice as many variables, and that the matrix which controls the fluctuations is now simple $M_{\rm ij} = JN_{\rm ij}$. The results are

$$\mathcal{Z}_{\rm dir}(J) \propto \exp\left(-\frac{d-1}{2}\sum_{\rm a=2}^{N} \ln[J\Lambda_{\rm a}]\right), \qquad \textbf{[S64]}$$

and

$$\langle \vec{\pi}_{\rm i} \cdot \vec{\pi}_{\rm j} \rangle = (d-1)\sum_{\rm a=2}^{N} \frac{w_{\rm i}^{\rm a} w_{\rm j}^{\rm a}}{J\Lambda_{\rm a}}, \qquad \textbf{[S65]}$$

where we give the result for motion in d dimensions; here $d = 3$.

As noted at the end of section II, the constraint that expectations values in our model equal those in the data is equivalent to maximum likelihood inference. Thus, to complete our calculation and find the parameters of our model, we should compute the probability of the data in the model as a function of the parameters J, g, and n_c. Putting together the results in this section, we can write the log of the full probability distribution as

$$\Phi \equiv \ln P(\text{data} \mid \text{model}) = -\ln \mathcal{Z} - \langle \mathcal{H}(\mathbf{v}) \rangle_{\rm exp} \qquad \textbf{[S66]}$$

$$= -\ln \mathcal{Z}_{\rm dir}(J) - \ln \mathcal{Z}_{\rm sp}(J,g) - \left\langle \frac{J}{4V^2}\sum_{\rm i,j=1}^{N} n_{\rm ij} |\vec{\mathbf{v}}_{\rm i} - \vec{\mathbf{v}}_{\rm j}|^2 \right\rangle_{\rm exp} - \left\langle \frac{g}{2V^2}\sum_{\rm i=1}^{N} (v_{\rm i} - V)^2 \right\rangle_{\rm exp} \qquad \textbf{[S67]}$$

$$= \sum_{\rm a=2}^{N} \ln[J\Lambda_{\rm a}] + \frac{1}{2}\sum_{\rm a=2}^{N} \ln[g + J\Lambda_{\rm a}] - N\frac{Jn_c}{2}\langle Q_{\rm int} \rangle_{\rm exp} - N\frac{g}{2}\langle \sigma^2 \rangle_{\rm exp}, \qquad \textbf{[S68]}$$

where $\langle \cdots \rangle$ denotes an average over the data. We have identified $Q_{\rm int}$ from Eq. **1** of the main text, and σ^2 is the fractional variance of individual birds' speeds around the flock mean, from Eq. **2** of the main text.

The result for Φ in Eq. **S68** is simple enough that we can maximize to give explicit equations that determine the parameters. Thus

$$\frac{\partial \Phi}{\partial g} = 0 \qquad \textbf{[S69]}$$

$$\Rightarrow \frac{1}{N}\sum_{a=2}^{N}\frac{1}{g+J\Lambda_a}=\langle\sigma^2\rangle_{exp}, \quad \text{[S70]}$$

and similarly

$$\frac{\partial\Phi}{\partial J}=0 \quad \text{[S71]}$$

$$\Rightarrow \frac{(N-1)}{J}+\frac{1}{2}\sum_{a=2}^{N}\frac{\Lambda_a}{g+J\Lambda_a}=N\frac{n_c}{2}\langle Q_{int}\rangle_{exp} \quad \text{[S72]}$$

$$d\left(1-\frac{1}{N}\right)-g\langle\sigma^2\rangle_{exp}=Jn_c\langle Q_{int}\rangle_{exp}. \quad \text{[S73]}$$

Finally, we can substitute the solutions to these equations, J^* and g^*, back into **Eq. S68** and maximize with respect to n_c, as in Fig. 2*C* of the main text.

B. Computation with Fixed Boundary Conditions. So far, we have assumed free boundary conditions, corresponding to the ideal situation where speed and orientations of all individuals in a flock can fluctuate in the same manner, exploring the whole accessible space of possible fluctuations, given the interaction between birds. In natural flocks this is not very realistic: individuals on the boundary are constantly subject to environmental stimuli, so that they will adjust their direction and speed not only in response to neighboring birds, but also in response to external cues. To cope with this fact, we now perform the computation of the partition functions and of the likelihood using fixed boundary conditions, where the velocities of the birds on the boundary of the flock are held fixed at their observed values. We note that for large systems, such as the flocks we are considering, boundary individuals are a negligible fraction of all individuals. As discussed more fully in ref. 7, the values of the inferred parameters do not change much when changing the boundary conditions. Fixed boundary conditions are however necessary to adequately take into account the effects of boundary on the correlations.

To perform the computations with fixed conditions on the border, it is convenient to divide the birds in two groups: internal birds $i,j\in\mathcal{I}$ and birds belonging to the border $a,b\in\mathcal{B}$. Then, Eq. **S27** becomes

$$\mathcal{H}(\mathbf{v})=\frac{J}{2v_0^2}\sum_{i,j\in\mathcal{I}}\left(N_{ij}+\frac{g}{J}\delta_{ij}\right)\vec{\mathbf{v}}_i\cdot\vec{\mathbf{v}}_j-\frac{J}{v_0}\sum_{i\in\mathcal{I}}\vec{\mathbf{h}}_i\cdot\vec{\mathbf{v}}_i+\mathcal{H}_{\mathcal{B}}(J,g)-\frac{\mu}{v_0}\sum_{i=1}^{N}v_i, \quad \text{[S74]}$$

where

$$\vec{\mathbf{h}}_i=\frac{1}{v_0}\sum_{a\in\mathcal{B}}n_{ia}\vec{\mathbf{v}}_a \quad \text{[S75]}$$

$$\mathcal{H}_{\mathcal{B}}(J,g)=\frac{J}{2v_0^2}\sum_{a,b\in\mathcal{B}}\left(N_{ab}+\frac{g}{J}\delta_{ab}\right)\vec{\mathbf{v}}_a\cdot\vec{\mathbf{v}}_b. \quad \text{[S76]}$$

We can see from these expressions that holding velocities $\vec{\mathbf{v}}_a$ fixed on the border of the flock is equivalent to considering a flock in presence of a field $\vec{\mathbf{h}}_i$ acting on those birds who see the border birds as their neighbors. Note that birds deep in the interior do not couple directly to the field, but may feel its influence if it propagates through the flock. It will be useful to decompose these fields in relation to the mean flight direction $\hat{\mathbf{n}}$, as in Eq. **S39**,

$$\vec{\mathbf{h}}_i=h_i^L\hat{\mathbf{n}}+\vec{\mathbf{h}}_i^{\perp}. \quad \text{[S77]}$$

The computation of the partition function now proceeds exactly as in subsection IV.A. The only difference is that integrations must now be performed on internal variables only; the algebra is slightly more complicated, but the conceptual steps are the same. Corresponding to Eq. **S55** we have

$$\mathcal{Z}(J,g;n_c)=e^{-\mathcal{H}_{\mathcal{B}}(J,g)}\mathcal{Z}_{dir}(J)\mathcal{Z}_{sp}(J,g)e^{Ng/2}, \quad \text{[S78]}$$

and in place of Eqs. **S56** and **S57** we have

$$\mathcal{Z}_{dir}(J)=\int\left[\prod_{i\in\mathcal{I}}d^2\boldsymbol{\pi}_i\right]\delta\left(\sum_{i=1}^{N}\vec{\boldsymbol{\pi}}_i\right)e^{-\mathcal{H}_{dir}(\{\vec{\boldsymbol{\pi}}_{i\in\mathcal{I}}\})} \quad \text{[S79]}$$

$$\mathcal{Z}_{sp}(J,g)=\int\left[\prod_{i\in\mathcal{I}}d\epsilon_i\right]\delta\left(\sum_{j=1}^{N}\epsilon_j\right)e^{-\mathcal{H}_{sp}(\{\epsilon_{i\in\mathcal{I}}\})}, \quad \text{[S80]}$$

where we note that the integration is only over internal variables, but the delta function constraints involve all of the variables. As in the case of free boundaries, we first integrate over global variables, which has the effect of pinning the mean velocity to its observed value, and then we can choose the scale $v_0=\langle V\rangle_{exp}$, simplifying all of the expressions. The reduced Hamiltonians for the internal variables, analogs of Eqs. **S53** and **S54**, then become

$$\mathcal{H}_{dir}(\{\vec{\boldsymbol{\pi}}_{i\in\mathcal{I}}\})=\frac{J}{2}\sum_{i,j\in\mathcal{I}}N_{ij}\vec{\boldsymbol{\pi}}_i\cdot\vec{\boldsymbol{\pi}}_j-J\sum_{i\in\mathcal{I}}\vec{\mathbf{h}}_i^{\perp}\cdot\vec{\boldsymbol{\pi}}_i \quad \text{[S81]}$$

$$\mathcal{H}_{sp}(\{\epsilon_{i\in\mathcal{I}}\})=\frac{J}{2}\sum_{i,j\in\mathcal{I}}\left(N_{ij}+\frac{g}{J}\delta_{ij}\right)\epsilon_i\epsilon_j-J\sum_{i\in\mathcal{I}}b_i\epsilon_i, \quad \text{[S82]}$$

where

$$b_i=h_i^L-\sum_{a\in\mathcal{B}}n_{ia}=\sum_{a\in\mathcal{B}}n_{ia}\epsilon_a \quad \text{[S83]}$$

is the fluctuating part of the longitudinal component of the border field.

Although we have the same matrix N_{ij} in these equations as in subsection IV.A, the indices ij are restricted to the interior of the flock, and on this restricted space the matrix has different properties. To remind us of this fact, it is convenient to introduce the two matrices $A_{ij}=N_{ij}$ and $B_{ij}=N_{ij}+(g/J)\delta_{ij}$, with indices that refer only to birds internal to the flock, $i\in\mathcal{I}$. Then the partition functions that we need to evaluate are again Gaussian integrals, controlled by the properties of these matrices. We find, corresponding to Eqs. **S61** and **S64**,

$$\begin{aligned}\ln\mathcal{Z}_{dir}(J)=&\frac{J}{2}\sum_{i,j\in\mathcal{I}}(A^{-1})_{ij}\vec{\mathbf{h}}_i^{\perp}\cdot\vec{\mathbf{h}}_j^{\perp}-\frac{d-1}{2}(N_{\mathcal{I}}-1)\ln(J)\\&-\frac{d-1}{2}\ln\left[\sum_{i,j\in\mathcal{I}}(A^{-1})_{ij}\right]-\frac{d-1}{2}\ln\det A\\&-\frac{J}{2}\left|\sum_{a\in\mathcal{B}}\vec{\boldsymbol{\pi}}_a+\sum_{i,j\in\mathcal{I}}(A^{-1})_{ij}\vec{\mathbf{h}}_i^{\perp}\right|^2\frac{1}{\sum_{i,j\in\mathcal{I}}(A^{-1})_{ij}}, \quad \text{[S84]}\end{aligned}$$

and

$$\ln \mathcal{Z}_{sp}(J,g) = \frac{J}{2}\sum_{i,j\in\mathcal{I}}(B^{-1})_{ij}b_i b_j - \frac{1}{2}(N_{\mathcal{I}}-1)\log(J) - \frac{1}{2}\log\left[\sum_{i,j\in\mathcal{I}}(B^{-1})_{ij}\right] - \frac{1}{2}\ln\det B - \frac{J}{2}\left|\sum_{a\in\mathcal{B}}\epsilon_a + \sum_{i,j\in\mathcal{I}}(B^{-1})_{ij}b_i\right|^2 \frac{1}{\sum_{i,j\in\mathcal{I}}(B^{-1})_{ij}}. \quad \textbf{[S85]}$$

Similarly, the probability distributions of the variables $\{\epsilon_i, \vec{\pi}_i\}$ again are Gaussian, and we can find, by analogy with Eqs. **S63** and **S65**, the correlation functions. One new feature is that birds in the interior can have nonzero averages of these fluctuations, since they are responding to the birds on the boundary. Instead of rotating to the basis of eigenvectors, it is useful to define the matrices

$$\tilde{A}_{ij} = (A^{-1})_{ij} - \frac{\sum_{l\in\mathcal{I}}(A^{-1})_{il}\sum_{m\in\mathcal{I}}(A^{-1})_{jm}}{\sum_{l,m\in\mathcal{I}}(A^{-1})_{lm}}, \quad \textbf{[S86]}$$

$$\tilde{B}_{ij} = (B^{-1})_{ij} - \frac{\sum_{l\in\mathcal{I}}(B^{-1})_{il}\sum_{m\in\mathcal{I}}(B^{-1})_{jm}}{\sum_{l,m\in\mathcal{I}}(B^{-1})_{lm}}. \quad \textbf{[S87]}$$

Then we find the mean directional fluctuation and the correlations in these fluctuations to be

$$\langle \vec{\pi}_i \cdot \vec{\pi}_j \rangle = \frac{d-1}{J}\tilde{A}_{ij} + \langle \vec{\pi}_i \rangle \cdot \langle \vec{\pi}_j \rangle, \quad \textbf{[S88]}$$

$$s\langle \vec{\pi}_i \rangle = \sum_{j\in\mathcal{I}} \tilde{A}_{ij}\vec{\mathbf{h}}_j^{\perp} - \frac{\sum_{j\in\mathcal{I}}(A^{-1})_{ij}}{\sum_{l,m\in\mathcal{I}}(A^{-1})_{lm}}\sum_{a\in\mathcal{B}}\vec{\pi}_a. \quad \textbf{[S89]}$$

Similarly, we find the mean speed fluctuation and correlations to be

$$\langle \epsilon_i \cdot \epsilon_j \rangle = \frac{1}{J}\tilde{B}_{ij} + \langle \epsilon_i \rangle \cdot \langle \epsilon_j \rangle, \quad \textbf{[S90]}$$

$$\langle \epsilon_i \rangle = \sum_{j\in\mathcal{I}} \tilde{B}_{ij} b_j - \frac{\sum_{j\in\mathcal{I}}(B^{-1})_{ij}}{\sum_{l,m\in\mathcal{I}}(B^{-1})_{lm}}\sum_{a\in\mathcal{B}}\epsilon_a. \quad \textbf{[S91]}$$

The correlation functions that we present in Figs. 2 and 3 of the main text are based on these expressions.

Finally, we need to find the conditions that set the values of the parameters. By analogy with Eqs. **S70** and **S73**, we find

$$\frac{1}{J}\left(d\frac{N^{\mathcal{I}}-1}{N} - g\langle\sigma^2\rangle_{exp}\right) = n_c\langle Q_{int}\rangle_{exp} + \frac{1}{N}\sum_{i,j\in\mathcal{I}}(A^{-1})_{ij}\vec{\mathbf{h}}_i^{\perp}\cdot\vec{\mathbf{h}}_j^{\perp} + \frac{1}{N}\sum_{i,j\in\mathcal{I}}(B^{-1})_{ij}b_i b_j - \frac{1}{N}\left|\sum_{a\in\mathcal{B}}\vec{\pi}_a + \sum_{i,j\in\mathcal{I}}(A^{-1})_{ij}\vec{\mathbf{h}}_j^{\perp}\right|^2\frac{1}{\sum_{i,j\in\mathcal{I}}(A^{-1})_{ij}} - \frac{1}{N}\left(\sum_{a\in\mathcal{B}}\epsilon_a + \sum_{i,j\in\mathcal{I}}(B^{-1})_{ij}b_j\right)^2\frac{1}{\sum_{i,j\in\mathcal{I}}(B^{-1})_{ij}} - \frac{1}{2N\langle V\rangle_{exp}^2}\sum_{a,b\in\mathcal{B}}n_{ab}|\vec{\mathbf{v}}_a - \vec{\mathbf{v}}_b|^2 + \frac{g}{NJ\langle V\rangle_{exp}^2}\sum_{a\in\mathcal{B}}\left(v_a - \langle V\rangle_{exp}\right)^2, \quad \textbf{[S92]}$$

and

$$\langle\sigma^2\rangle_{exp} = \frac{1}{NJ}\sum_{i\in\mathcal{I}}\tilde{B}_{ii} + \frac{1}{N}\sum_{a\in\mathcal{B}}\epsilon_a^2 + \frac{1}{N}\frac{\sum_{i,j\in\mathcal{I}}(B^{-2})_{ij}}{\sum_{i,j\in\mathcal{I}}(B^{-1})_{ij}}\left(\sum_{a\in\mathcal{B}}\epsilon_a\right)^2. \quad \textbf{[S93]}$$

Finally, the optimal value of n_c can be found by maximizing the log-likelihood

$$\Phi(J,g;n_c) = -\ln\mathcal{Z}_{dir}(J) - \ln\mathcal{Z}_{sp}(J,g) + \mathcal{H}_{\mathcal{B}}(J,g) - \frac{Jn_cN}{2}\langle Q_{int}\rangle_{exp} - N\frac{g}{2}\langle\sigma^2\rangle_{exp}, \quad \textbf{[S94]}$$

where we substitute for J and g the (n_c-dependent) solutions of Eqs. **S92** and **S93**. An example of the likelihood as a function of n_c is given in the main text.

V. Summary of Model Parameters

Eqs. **S88–S94** give us all of the ingredients needed to determine the parameters J, g, and n_c, as in Fig. 1 of the main text. A summary of these results is given in Table S2.

As in our previous work (7), we find that parameters do not vary dramatically from flock to flock. In particular, as shown in Fig. S1, our estimate of the neighborhood size n_c does not vary systematically with the linear dimensions of the flock. This is important because we want to be sure that our account of scale-free correlations does not hinge on adjusting the range of interactions to the size of the flock.

The next important point is that the neighborhood size is independent of the typical distance between birds (Fig. S2). This reinforces the conclusion (4, 7) that birds interact with a fixed number of neighbors, independent of distance—a "topological" rather than metric interaction.

Most important for our present discussion is the consistently small value of the ratio $g/(Jn_c)$, which is a sign of the approach to criticality. In fact we can even see that $g/(Jn_c)$ is slightly negative, which is at first sight surprising. In fact, from Eq. **S63** we see that what must be positive is not g, but rather $g + J\Lambda_2$. The lowest nonzero eigenvalue Λ_2 becomes smaller as the flock becomes larger, so that only modestly negative values of g are allowed. There is a correction to this picture when the birds on the border are fixed, but the idea is the same: g can be slightly negative if this is compensated by other sources of stiffness.

VI. Goldstone Modes and the Continuum Limit

In this section we would like to make some of the mathematics behind the intuitions described in *Some Intuition* of the main text more explicit. Our discussion is for the case (section IV, A) with free boundary conditions.

We start by looking at the effective Hamiltonian for the directional variables $\{\vec{\pi}_i\}$, in Eq. **S53**,

$$\mathcal{H}_{dir}(\{\vec{\pi}_i\}) = \frac{J}{2}\sum_{i,j=1}^{N}N_{ij}\vec{\pi}_i\cdot\vec{\pi}_j.$$

As explained in the discussion leading up Eq. **S58**, the matrix N_{ij} has a zero eigenvalue, but in fact the whole eigenvalue spectrum has a special structure. To see this, it is useful to imagine that the birds are arranged along a line, and that the neighborhood is only the very nearest neighbor. Then we can label the birds by n, and the bird $n+1$ is the neighbor of bird n; we can rearrange the terms in the sum to give

$$\mathcal{H}_{\rm dir}\left(\{\vec{\pi}_{\rm i}\}\right)=\frac{J}{2}\sum_{\rm n=1}^{N}|\vec{\pi}_{\rm n}-\vec{\pi}_{n+1}|^2. \qquad \textbf{[S95]}$$

Now suppose that the direction of flight varies only very slowly, so that we can picture a continuous function of position x in the flock, despite the fact that the birds are located at discrete positions $x_{\rm n}=nr_c$, where r_c is the typical distance between the nearest birds. Then we have $\vec{\pi}(x)$, and

$$\mathcal{H}_{\rm dir}\left(\{\vec{\pi}_{\rm i}\}\right)\approx\frac{Jr_c^2}{2}\sum_{\rm n=1}^{N}\left|\frac{\partial\vec{\pi}(x)}{\partial x}\right|^2. \qquad \textbf{[S96]}$$

Since we are assuming that variations are smooth, we can turn the sum into an integral,

$$\mathcal{H}_{\rm dir}\left(\{\vec{\pi}_{\rm i}\}\right)=\frac{Jr_c^2}{2}\rho\int dx\left|\frac{\partial\vec{\pi}(x)}{\partial x}\right|^2, \qquad \textbf{[S97]}$$

where ρ is the density of birds along the line. If we do the same calculation with birds on a regular lattice in 3D space rather than birds along a line, we find

$$\mathcal{H}_{\rm dir}\left(\{\vec{\pi}_{\rm i}\}\right)=\frac{Jn_cr_c^2}{2}\rho\int d^3x|\nabla\vec{\pi}(x)|^2, \qquad \textbf{[S98]}$$

where we also include the more realistic possibility that the neighborhood is not just one neighbor but a group of n_c neighbors.

The crucial point about Eq. **S98** is that if we consider variations in flight direction on a scale ℓ, such as $\vec{\pi}(x)\sim A\sin(2\pi x/\ell)$, then we have $\mathcal{H}_{\rm dir}\propto A^2/\ell^2$. Thus, as the length scale of variations becomes large ($\ell\to\infty$), the stiffness which resists the variations goes to zero. This vanishing stiffness at long wavelengths is the signature of a Goldstone mode, which arises because the original model allowed flight in any direction, but the actual state of the flock breaks this symmetry by selecting a particular direction (13).

If the stiffness that opposes variations (in the Hamiltonian) goes down, then the variance of these fluctuations (in the probability distribution) goes up. Thus, in the presence of Goldstone modes we will see a large variance of fluctuations corresponding to variations over long-length scales. In other words, we will see long-ranged correlations. It is important these correlations not only be long ranged, but also genuinely scale-free. To see this, it is useful to remember several mathematical facts about Gaussian random functions (see, for example, appendix A.2 in ref. 8).

Suppose that we have a function $\phi(x)$, with zero mean. If all points x are equivalent, we can characterize the statistics of fluctuations in $\phi(x)$ using the correlation function,

$$C_\phi(x-x')=\langle\phi(x)\phi(x')\rangle. \qquad \textbf{[S99]}$$

It is also useful to consider the Fourier transform of the correlation function, the power spectrum,

$$S_\phi(k)=\int dx\,e^{+ikx}C_\phi(x). \qquad \textbf{[S100]}$$

Importantly, we can write the entire probability distribution for the functions $\phi(x)$ using the power spectrum,

$$P[\phi(x)]=\frac{1}{Z}\exp\left[-\frac{1}{2}\int\frac{dk}{2\pi}\frac{|\tilde{\phi}(k)|^2}{S_\phi(k)}\right], \qquad \textbf{[S101]}$$

where

$$\tilde{\phi}(k)=\int dx\,e^{+ikx}\phi(x) \qquad \textbf{[S102]}$$

is the Fourier transform of the function $\phi(x)$.

Since we have $P\propto\exp[-\mathcal{H}]$, Eq. **S98** tells us that

$$P\left[\vec{\pi}(x)\right]=\frac{1}{Z}\exp\left[-\frac{Jn_cr_c^2}{2}\rho\int d^3x\left|\nabla\vec{\pi}(\vec{\mathbf{x}})\right|^2\right]. \qquad \textbf{[S103]}$$

We can also write this in terms of the Fourier transforms,

$$\vec{\pi}(\vec{\mathbf{k}})=\int dx\,e^{+i\vec{\mathbf{k}}\cdot\vec{\mathbf{x}}}\vec{\pi}(\vec{\mathbf{x}}), \qquad \textbf{[S104]}$$

and then Eq. **S103** becomes

$$P\left[\vec{\pi}(x)\right]=\frac{1}{Z}\exp\left[-\frac{Jn_cr_c^2}{2}\rho\int\frac{d^3k}{(2\pi)^3}\left|\vec{\mathbf{k}}\right|^2\left|\tilde{\pi}(\vec{\mathbf{k}})\right|^2\right]. \qquad \textbf{[S105]}$$

However, now we can read off the power spectrum, by comparing Eqs. **S105** and **S101**; we see that

$$S_\pi(\vec{\mathbf{k}})=\frac{1}{Jn_cr_c^2\rho}\cdot\frac{1}{\left|\vec{\mathbf{k}}\right|^2}. \qquad \textbf{[S106]}$$

If we transform back to give the correlation function, we have

$$C_\pi(\vec{\mathbf{x}})=\int\frac{d^3k}{(2\pi)^3}e^{-i\vec{\mathbf{k}}\cdot\vec{\mathbf{x}}}S_\pi(\vec{\mathbf{k}}) \qquad \textbf{[S107]}$$

$$=\frac{1}{Jn_cr_c^2\rho}\int\frac{d^3k}{(2\pi)^3}e^{-i\vec{\mathbf{k}}\cdot\vec{\mathbf{x}}}\frac{1}{\left|\vec{\mathbf{k}}\right|^2}. \qquad \textbf{[S108]}$$

The key point about this result is that there is nothing in the integral to set a characteristic scale for $\vec{\mathbf{x}}$. In fact, if we double the value of $|\vec{\mathbf{x}}|$ we make up for this by cutting the value of $|\vec{\mathbf{k}}|$ in half so that $\vec{\mathbf{k}}\cdot\vec{\mathbf{x}}$ stays fixed, but since we are integrating over all possible values of $\vec{\mathbf{k}}$, only the whole integral is reduced by a factor of 2. This dimensional analysis argument tells us that

$$C_\pi(\vec{\mathbf{x}})\propto\frac{1}{|\vec{\mathbf{x}}|}. \qquad \textbf{[S109]}$$

This is a power-law decay of correlations with distance (here the power is 1), and it has no characteristic scale. Thus, scale-free correlations in directional fluctuations are a consequence of the Goldstone modes.

The predictions for speed fluctuations are very different from those for directional fluctuations. In taking the limit of smooth, continuous variations for directional variations, we found

$$\mathcal{H}_{\rm dir}\left(\{\vec{\pi}_{\rm i}\}\right)=\frac{J}{2}\sum_{\rm i,j=1}^{N}N_{\rm ij}\vec{\pi}_{\rm i}\cdot\vec{\pi}_{\rm j}\to\frac{Jn_cr_c^2}{2}\rho\int d^3x|\nabla\vec{\pi}(x)|^2. \qquad \textbf{[S110]}$$

The same argument for speed fluctuations starts with **Eq. S49**, and gives

$$\mathcal{H}_{\rm sp}(\{\epsilon_{\rm i}\}) = \frac{1}{2}\sum_{\rm i,j=1}^{N}\left(\frac{gV^2}{v_0^2}\delta_{\rm ij} + JN_{\rm ij}\right)\epsilon_{\rm i}\epsilon_{\rm j} \qquad \text{[S111]}$$

$$\to \frac{1}{2}\rho\int d^3x\left[Jn_c r_c^2\left|\nabla\epsilon(\vec{\mathbf{x}})\right|^2 + g\epsilon^2(\vec{\mathbf{x}})\right]$$

$$= \frac{1}{2}\rho\int\frac{d^3k}{(2\pi)^3}\left[Jn_c r_c^2\left|\vec{\mathbf{k}}\right|^2 + g\right]\left|\tilde{\epsilon}(\vec{\mathbf{k}})\right|^2, \qquad \text{[S112]}$$

where in the last step we transform to the Fourier representation. With the same argument that leads to Eq. **S106**, we recognize the predicted power spectrum for fluctuations in the speed,

$$S_\epsilon(\vec{\mathbf{k}}) = \frac{1}{Jn_c r_c^2\rho}\cdot\frac{1}{\left|\vec{\mathbf{k}}\right|^2 + g/(Jn_c r_c^2)}. \qquad \text{[S113]}$$

Thus, where $S_{\vec{\pi}}$ grows without bound as the wave vector $\vec{\mathbf{k}}$ becomes small, $S_\epsilon(\vec{\mathbf{k}})$ stops growing once $\vec{\mathbf{k}}$ is smaller than a characteristic scale $k_c = 1/\xi = \sqrt{g/(Jn_c r_c^2)}$. We note that ξ is a length, and we expect that this defines the distance across which significant correlations can be observed. Indeed, if we transform back to get the correlation function, we have

$$C_\epsilon(\vec{\mathbf{x}}) = \int\frac{d^3k}{(2\pi)^3}e^{-i\vec{\mathbf{k}}\cdot\vec{\mathbf{x}}}S_\epsilon(\vec{\mathbf{k}}) \qquad \text{[S114]}$$

$$= \frac{1}{Jn_c r_c^2\rho}\int\frac{d^3k}{(2\pi)^3}e^{-i\vec{\mathbf{k}}\cdot\vec{\mathbf{x}}}\frac{1}{\left|\vec{\mathbf{k}}\right|^2 + g/(Jn_c r_c^2)} \qquad \text{[S115]}$$

$$\propto e^{-|\vec{\mathbf{x}}|/\xi}, \qquad \text{[S116]}$$

corresponding to Eq. **10** of the main text.

From these results we can see that, for generic values of g/J, the maximum entropy model predicts very different kinds of correlations for directions and speeds. In the case of directions, the correlations have a dominant contribution from long wavelength modes, there is no intrinsic length scale, and we see scale-free behavior. On the contrary, in the case of speed fluctuations the contribution of the long wavelength modes is cut off by the term g/J, which by analogy with field theory we can think of as a mass (13); this results in correlations that decay exponentially with the distance between birds. However, when g/J goes to zero, or, more precisely, when the predicted correlation length ξ becomes comparable to the linear dimensions of the flock as whole, our analysis breaks down. We have described an essentially infinite system, with no boundaries. When g/Jn_c is small enough that $\xi \sim r_c\sqrt{Jn_c/g} \sim L$, then the whole flock is effectively correlated, and a more detailed analysis is needed. We shall see that, in this critical regime, it is possible for the speed fluctuations also to be scale-free.

VII. Decoupling Speeds and Flight Directions

The approach we have taken thus far is to build the least structured models that are consistent with the observed similarity of velocities between birds and their near neighbors. Importantly, we treat the velocities as vectors, and use a measure of similarity that is a rotationally invariant, analytic function of these vectors, $Q_{\rm int}$ in Eq. **1** in the main text and Eq. **S20** above. One could imagine, however, that real birds do not obey these symmetries. In particular, they could have very separate mechanisms for adjusting their speeds and directions in relation to those of their neighbors, or their perceptual apparatus for estimating speeds and directions may introduce errors that are not equivalent to an isotropic vector error. Under these conditions, it would make more sense to build models that have separate constraints for the observed degree of speed and direction similarity among neighbors, and this is what we explore here below.

We can measure the degree of similarity or correlation among directions in the same way that we did in ref. 7, defining

$$C_{\rm int} = \frac{1}{N}\sum_{\rm i=1}^{N}\frac{1}{n_c^{\rm dir}}\sum_{\rm j\in\mathcal{N}_i^{dir}}\vec{\mathbf{s}}_{\rm i}\cdot\vec{\mathbf{s}}_{\rm j}, \qquad \text{[S117]}$$

where we allow that the neighborhood for measuring directional similarity may have a size $n_c^{\rm dir}$ that differs from the corresponding neighborhood for measuring speed similarity, $n_c^{\rm sp}$. We can also define a (dis)similarity measure for the speeds, by analogy with $Q_{\rm int}$,

$$Q_{\rm int}^{\rm sp} = \frac{1}{2Nv_0^2}\sum_{\rm i=1}^{N}\frac{1}{n_c^{\rm sp}}\sum_{\rm j\in\mathcal{N}_i^{sp}}\left(v_{\rm i} - v_{\rm j}\right)^2. \qquad \text{[S118]}$$

If we build the maximum entropy model consistent with measured values of these quantities, plus the mean and variance of individual speeds across the flock, we obtain, instead of Eq. **S27**,

$$\mathcal{H}(\mathbf{v}) = \frac{J^{\rm sp}}{4v_0^2}\sum_{\rm i,j=1}^{N}n_{\rm ij}^{\rm sp}\left(v_{\rm i} - v_{\rm j}\right)^2 - \frac{J^{\rm dir}}{2v_0^2}\sum_{\rm i,j=1}^{N}n_{\rm ij}^{\rm dir}\vec{\mathbf{s}}_{\rm i}\cdot\vec{\mathbf{s}}_{\rm j} + \frac{g}{2v_0^2}\sum_{\rm i=1}^{N}v_{\rm i}^2 - \frac{\mu}{v_0}\sum_{\rm i=1}^{N}v_{\rm i}, \qquad \text{[S119]}$$

where $n_{\rm ij}^{\rm sp}$ is defined as $n_{\rm ij}$ above, but with neighborhoods of size $n_c^{\rm sp}$, and similarly for $n_{\rm ij}^{\rm dir}$. Notice that we now have two different coupling strengths, $J^{\rm sp}$ and $J^{\rm dir}$, controlling speed and directional ordering, respectively.

Because our original model breaks into separate pieces for directional and speed fluctuations, we can carry over all of the calculations, being careful about the values of the parameters. If we set $J^{\rm sp} = J^{\rm dir}$ we are back to our original model. With the two separate parameters we find the log-likelihood, by analogy with Eq. **S94**,

$$\Phi = -\ln\mathcal{Z}_{\rm dir}(J^{\rm dir}; n_c^{\rm dir}) - \ln\mathcal{Z}_{\rm sp}(J^{\rm sp}, g; n_c^{\rm sp}) + \mathcal{H}_{\mathcal{B}} - \frac{J^{\rm sp}n_c^{\rm sp}N}{2}\langle Q_{\rm int}^{\rm sp}\rangle_{\rm exp} + \frac{J^{\rm dir}n_c^{\rm dir}N}{2}\langle C_{\rm int}\rangle_{\rm exp} - N\frac{g}{2}\langle\sigma^2\rangle_{\rm exp}. \qquad \text{[S120]}$$

We can then infer, independently for speed and orientation, the interaction parameters, and compare them to see how different they are. We can also check whether and how much the predictions for the correlation functions are better than in the simpler, unified model. Results are shown in Fig. S3. We can see that for most flocks the global interaction strength Jn_c for the speed and directional degrees of freedom are very similar to each other (Fig. S3*A*): in this case the unified model discussed earlier is basically equivalent to this more general model, both in terms of values of inferred parameters and in terms of predictions for the correlation functions. For a few flocks, however, we observe a decoupling between flight directions and speeds. This typically occurs when the fractional speed fluctuations are on a different scale from the directional fluctuations. In these cases, the model that fixes the local similarities of speed and direction separately provides better predictions for the speed correlations than the unified model (Fig. S3*B*), although these differences are not large.

Building a model that fixes the local similarities of speed and direction separately must provide a more accurate description of the system, since it imposes two different ways in which our model distribution $P(\mathbf{v})$ has to match the real distribution of (vector) velocities. The fact that the gain in accuracy usually is small seems significant, and suggests that those rare instances where differences are larger should have biological meaning. Indeed, in most of the events where the decoupling is stronger (on the right in Fig. S3*A*) the flocks are turning. Recent findings (14) show that additional conservation laws must be taken into account to explain the dynamics during the turn. Even if such conservation laws do not modify the form of the probability distribution we are investigating in the present work, they might give rise to different effective parameters for directions and speeds.

VIII. Dynamical Model

In this section we describe the dynamical model introduced in Eqs. **16** and **17** of the main text and its numerical implementation in more detail. We have

$$\gamma\frac{d\vec{\mathbf{v}}_i(t)}{dt} = -\nabla_i\mathcal{H}\left(\{\vec{\mathbf{v}}_j\}\right) + \vec{\eta}_i(t) \qquad \textbf{[S121]}$$

$$= -\frac{J}{2v_0^2}\sum_j n_{ij}\left(\vec{\mathbf{v}}_i - \vec{\mathbf{v}}_j\right) - \frac{g}{v_0^2}\frac{\vec{\mathbf{v}}_i}{v_i}\left(v_i - \hat{v}\right) + \frac{1}{n_c}\sum_{j\in\mathcal{N}_i}\vec{\mathbf{f}}_{ij} + \vec{\eta}_i(t) \qquad \textbf{[S122]}$$

$$\frac{d\vec{\mathbf{x}}_i}{dt} = \vec{\mathbf{v}}_i, \qquad \textbf{[S123]}$$

where we have added, as described in the text, forces $\vec{\mathbf{f}}_{ij}$ that serve to hold the flock together. If we write the vector components of $\vec{\eta}_i(t)$ as $\eta_i^\nu(t)$, with $\nu = 1, 2, 3$, then

$$\left\langle \eta_i^\nu(t)\eta_j^\mu(t')\right\rangle = 2\gamma T\delta_{ij}\delta_{\mu\nu}\delta(t-t'), \qquad \textbf{[S124]}$$

where T is an effective temperature for the noisy dynamics. We can chose our units of time so that $\gamma = 1$, and from the discussion in section IV, we can chose $\hat{v} = v_0 = \langle V\rangle_{\text{exp}}$, the desired mean speed of the flock.

In this form, the model that we are considering describes self-propelled particles (SPPs), and is very similar to the Vicsek model with attraction, which has been studied extensively in the literature (15–18). An attraction term is required to keep the flock cohesive in open space and prevent fluctuations and/or perturbations that disrupt the group. It has been shown that these effects are remarkably less important in models with topological interactions (4, 18, 19), which are much more robust in cohesion than SPP models with metric interactions. Nevertheless, even in the topological case, an attraction force is the most controlled way to fix the density of the group to a stationary value, therefore we will include it. We choose the forces

$$\vec{\mathbf{f}}_{ij} = \alpha\frac{\vec{\mathbf{r}}_{ij}}{r_{ij}}\begin{cases}\frac{1}{4}\frac{r_{ij}-r_e}{r_a-r_{hc}} & \text{if} \quad r_{ij} < r_a\\ 1 & \text{otherwise,}\end{cases} \qquad \textbf{[S125]}$$

where $\vec{\mathbf{r}}_{ij}$ is the vector from bird i to bird j, $r_{ij} = |\vec{\mathbf{r}}_{ij}|$ is its length, r_e is the equilibrium distance between birds where the force vanishes, while r_a and r_{hc} set spatial scales for the extent of the force. In our simulations we choose $r_e = 0.5$, $r_a = 0.8$, and $r_{hc} = 0.2$, which sets our units of length, and $\alpha = 0.95$.

An important point is that, when we sum the contributions of the forces $\vec{\mathbf{f}}_{ij}$, we include only birds within a limited neighborhood, $j\in\mathcal{N}_i$. As in the measure of similarity Q_{int}, this neighborhood is defined topologically, so that each bird feels the effect of n_c closest neighbors, rather than all of the birds within a fixed physical distance. In addition, for these simulations we introduced a balancing criterion, according to which a bird considers interacting neighbors homogeneously around it to coordinate with. This mimics the idea of a shell of relevant topological neighbors, and is similar to using Voronoi neighbors, as in ref. 19, but is much easier to implement numerically. A balanced interaction enhances the stability of the flock (18), increasing the range of parameters where Eqs. **S122** and **S123** give rise to realistic behavior. However, we checked also the simple topological case, obtaining qualitatively similar results.

Despite its similarity with other SPP models, the model we are considering has a crucial new ingredient, namely that the speeds of the individual birds are not fixed but can change in time. Accordingly, Eq. **S122** describes the evolution of the full velocity (rather than the flight direction, as in ref. 15), with a term $\propto g$ that sets the scale of the speed fluctuations. In addition, existing SPP models are usually defined as discrete dynamical update equations, which do not have a well-defined continuum limit. In contrast, we have defined our model as a stochastic differential equation.

We simulate our model using a finite interval (Euler) discretization, and we checked that macroscopic properties of the flock (e.g., the mean speed) remained the same if the size of the time step was decreased. Parameters J and n_c can be taken from the discussion of real flocks, and the temperature T adjusted until the polarization is in the range seen in the data (Table S1). We simulated flocks of different sizes, and checked that the flock had come to a stationary state before taking measurements. With all other parameters fixed, we varied g, with the results shown in Fig. 4 of the main text.

Long-ranged correlations can arise through one other mechanism that we have not discussed, and this is the emergence of hydrodynamic modes; it has been argued that such modes are an essential feature of SPP models on the largest spatial and temporal scales (20, 21). The simulations described here suggest, however, that such effects become dominant only on much larger scales in space and especially in time, and thus cannot explain the scale-free speed correlations that we observe at equal times in real flocks. We know that both metric and topological SPP/Vicsek models exhibit giant density fluctuations on large scales (19), yet we have seen that as long as g is finite, speed correlations are short range and a critical value of g is necessary to make them scale-free.

1. Cavagna A, et al. (2008) The STARFLAG handbook on collective animal behaviour: 1. Empirical methods. *Anim Behav* 76:217–236.
2. Cavagna A, Giardina I, Orlandi A, Parisi G, Procaccini A (2008) The STARFLAG handbook on collective animal behaviour: 2. Three-dimensional analysis. *Anim Behav* 76:237–248.
3. Cavagna A, et al. (2010) Scale-free correlations in starling flocks. *Proc Natl Acad Sci USA* 107(26):11865–11870.
4. Ballerini M, et al. (2008) Interaction ruling animal collective behavior depends on topological rather than metric distance: Evidence from a field study. *Proc Natl Acad Sci USA* 105(4):1232–1237.
5. Edelsbrunner H, Mücke EP (1994) Three-dimensional alpha shapes. *ACM Trans Graph* 13:43–72.
6. Jaynes ET (1957) Information theory and statistical mechanics. *Phys Rev* 106:620–630.
7. Bialek W, et al. (2012) Statistical mechanics for natural flocks of birds. *Proc Natl Acad Sci USA* 109(13):4786–4791.
8. Bialek W (2012) *Biophysics: Searching for Principles* (Princeton Univ Press, Princeton).
9. Shannon CE (1948) A mathematical theory of communication. *Bell Sys Tech J* 27(3): 379–423 and 27(4):623–656; reprinted in Shannon CE, Weaver W (1949) *The Mathematical Theory of Communication* (Univ of Illinois Press, Urbana, IL).
10. Cover TM, Thomas JA (1991) *Elements of Information Theory* (Wiley, New York).
11. Bender CM, Orszag SA (1978) *Advanced Mathematical Methods for Scientists and Engineers* (McGraw-Hill, New York).
12. Dyson FJ (1956) General theory of spin-wave interactions. *Phys Rev* 102(5):1217–1230.
13. Parisi G (1988) *Statistical Field Theory* (Addison-Wesley, Redwood City, CA).

14. Attanasi A, et al. (2013) Superfluid transport of information in turning of flocks, arXiv:1303.7097 [cond-mat.stat-mech].
15. Vicsek T, Czirók A, Ben-Jacob E, Cohen I, Shochet O (1995) Novel type of phase transition in a system of self-driven particles. *Phys Rev Lett* 75(6):1226–1229.
16. Grégoire G, Chaté H, Tu Y (2003) Moving and staying together without a leader. *Physica D* 181(3-4):157–170.
17. Grégoire G, Chaté H (2004) Onset of collective and cohesive motion. *Phys Rev Lett* 92(2):025702.
18. Camperi M, Cavagna A, Giardina I, Parisi G, Silvestri E (2012) Spatially balanced topological interaction grants optimal cohesion in flocking models. *Interface Focus* 2(6):715–725.
19. Ginelli F, Chaté H (2010) Relevance of metric-free interactions in flocking phenomena. *Phys Rev Lett* 105(16):168103.
20. Toner J, Tu Y (1995) Long-range order in a two-dimensional XY model: How birds fly together. *Phys Rev Lett* 75(23):4326–4329.
21. Toner J, Tu Y (1998) Flocks, herds, and schools: A quantitative theory of flocking. *Phys Rev E Stat Phys Plasmas Fluids Relat Interdiscip Topics* 58:4828–4858.

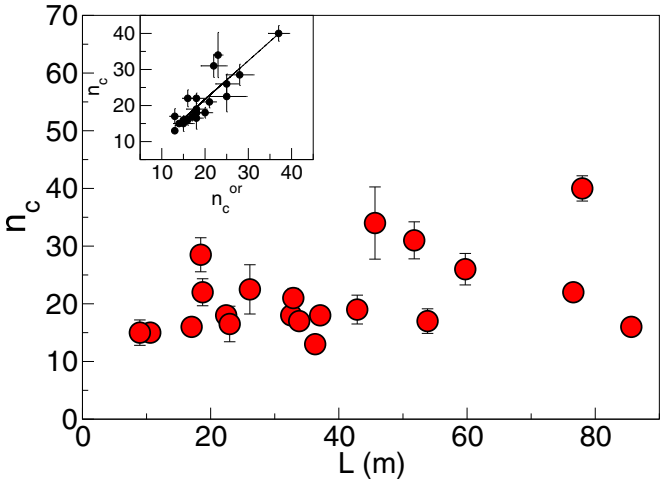

Fig. S1. Inferred values of n_c for flocks of different linear dimension. *Inset* shows that n_c determined here agrees with earlier results for a model of directional ordering alone (7). Error bars indicate SDs across multiple snapshots in each flocking event.

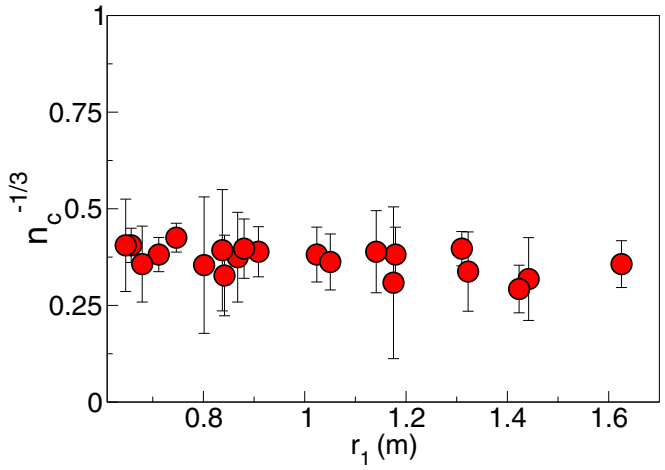

Fig. S2. Inferred values of n_c for flocks as a function of the typical distance r_1 between neighboring birds. Error bars same as in Fig. S1.

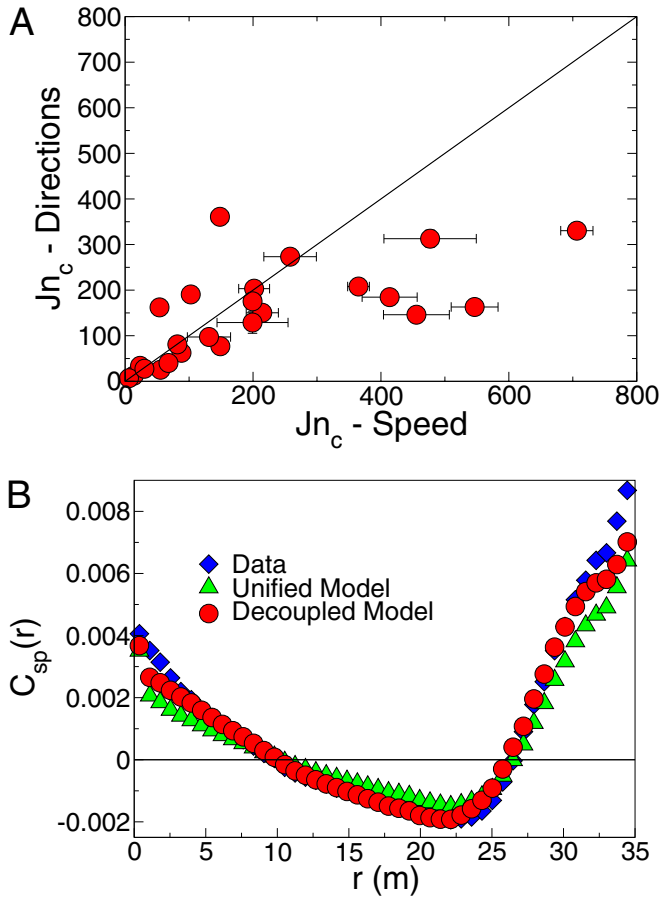

Fig. S3. Model with independent interactions for speed and flight directions. (*A*) The inferred global interaction strength Jn_c for the orientational degrees of freedom (y axis) vs. the speed degrees of freedom (x axis). The straight line corresponds to $y = x$, i.e., to the global model where the interaction parameters are the same for speed and flight directions. (*B*) Prediction for the speed correlation function of the unified model Eq. **S27** and for the decoupled model based on Eq. **S119**, for the flock corresponding to the rightmost point in *A* (28-10 in Tables S1 and S2).

Table S1. Summary of experimental data

Event	N	Frames	P	$\langle V\rangle_{exp}$, m/s	L, m	Q_{int}
17–06	552	79	0.935	9.96	51.8	1.29e-01
21–06	717	37	0.973	12.06	32.1	1.22e-02
25–08	1,571	29	0.962	12.47	59.8	2.63e-02
25–10	1,047	27	0.991	12.57	33.5	8.36e-03
25–11	1,176	30	0.959	10.07	43.3	6.27e-02
28–10	1,246	14	0.982	11.22	36.5	6.43e-03
29–03	440	27	0.963	10.75	37.1	1.43e-02
31–01	2,126	19	0.844	8.13	76.8	5.50e-02
32–06	809	34	0.981	9.99	22.2	1.52e-02
42–03	431	58	0.979	10.68	29.9	1.62e-02
49–05	797	16	0.995	14.02	19.2	6.49e-03
54–08	4,268	28	0.966	19.17	78.7	4.29e-02
57–03	3,242	16	0.978	14.38	85.7	1.53e-02
58–06	442	16	0.984	10.13	23.1	1.34e-02
58–07	554	14	0.977	10.81	19.1	1.35e-02
63–05	890	24	0.978	10.24	52.9	1.86e-02
69–09	239	56	0.985	11.97	17.1	2.68e-02
69–10	1,129	53	0.987	12.04	47.3	2.35e-02
69–19	803	18	0.975	14.16	26.4	3.65e-02
72–02	122	57	0.992	13.24	10.6	1.12e-02
77–07	186	22	0.978	9.50	9.1	4.27e-02

Flocking events are labeled according to experimental session number and to the position within the session to which they belong. The number of birds N is the number of individuals for which we obtained a 3D reconstruction of positions in space. The number of frames is the number of consecutive 3D reconstructions of individual positions and velocities in the flocking event at time intervals of 1/10 s. The polarization P is the global degree of alignment, as defined in the text. The linear size L of the flock is defined as the maximum distance between two birds belonging to the flock. The speed $\langle V\rangle_{exp}$ is the average of the individual speeds over all of the individuals in the flock, and Q_{int} is as defined in Eq. **1** of the main text. All values are averaged over several snapshots during the flocking event.

Table S2. Inferred parameters

Event	n_c	Jn_c	$g/(Jn_c)$
17–06	31	32.6	−4.1e-03
21–06	18	193.36	2.4e-02
25–08	26	71.76	2.0e-01
25–10	17	230.93	8.5e-02
25–11	19	135.95	−1.6e-02
28–10	13	361.32	1.5e-02
29–03	18	166.49	6.2e-01
31–01	22	80.07	2.9e-02
32–06	18	254.34	7.6e-02
42–03	21	150.16	2.4e-02
49–05	22	286.77	−7.1e-02
54–08	40	26.5	6.4e-01
57–03	16	89.20	9.5e-02
58–06	16.5	324.14	−8e-04
58–07	28.5	147.93	8.4e-01
63–05	17	155.6	−5e-02
69–09	16	86.74	2.6e-01
69–10	34	138.56	−1.2e-01
69–19	22.5	42.17	7.0e-02
72–02	15	179.88	2.5e-01
77–07	15	105.12	9.0e-02

Given a flocking event, we infer the values of the parameters J, g, and n_c using maximum entropy equations with fixed boundary conditions for each individual frame. We report in this table the median values (computed across all frames in a single event) of n_c, the global alignment strength Jn_c, and the effective control parameter $g/(Jn_c)$ that regulates the balance between adaptation and individual speed control.