Social interactions dominate speed control in poising natural flocks near criticality

William Bialek[a,1], Andrea Cavagna[b,c], Irene Giardina[b,c], Thierry Mora[d], Oliver Pohl[b,c,2], Edmondo Silvestri[b,c], Massimiliano Viale[b,c], and Aleksandra M. Walczak[e]

[a]Joseph Henry Laboratories of Physics and Lewis–Sigler Institute for Integrative Genomics, Princeton University, Princeton, NJ 08544; [b]Unità Operativa di Supporto (UOS) Sapienza, Istituto Sistemi Complessi–Consiglio Nazionale delle Ricerche, 00185 Rome, Italy; [c]Dipartimento di Fisica, Sapienza Universitá di Roma, 00185 Rome, Italy; [d]Laboratoire de Physique Statistique de l'École Normale Supérieure, Centre National de la Recherche Scientifique and Universités Paris VI and Paris VII, 75231 Paris Cedex 05, France; and [e]Laboratoire de Physique Théorique de l'École Normale Supérieure, Centre National de la Recherche Scientifique and Université Paris VI, 75231 Paris Cedex 05, France

Flocks of birds exhibit a remarkable degree of coordination and collective response. It is not just that thousands of individuals fly, on average, in the same direction and at the same speed, but that even the fluctuations around the mean velocity are correlated over long distances. Quantitative measurements on flocks of starlings, in particular, show that these fluctuations are scale-free, with effective correlation lengths proportional to the linear size of the flock. Here we construct models for the joint distribution of velocities in the flock that reproduce the observed local correlations between individuals and their neighbors, as well as the variance of flight speeds across individuals, but otherwise have as little structure as possible. These minimally structured or maximum entropy models provide quantitative, parameter-free predictions for the spread of correlations throughout the flock, and these are in excellent agreement with the data. These models are mathematically equivalent to statistical physics models for ordering in magnets, and the correct prediction of scale-free correlations arises because the parameters— completely determined by the data—are in the critical regime. In biological terms, criticality allows the flock to achieve maximal correlation across long distances with limited speed fluctuations.

collective behavior | statistical mechanics

In a flock of birds, thousands of individuals will fly in the same direction and at the same speed, for long periods of time. However, this average behavior is not enough for flocking to be advantageous. The entire flock must respond to dangers that may be visible only to a small fraction of individuals, requiring information to propagate over long distances. Although it is difficult to measure this information flow directly (1), we know that attacks by predators on a flock have very low success rates (2–4), and that the evasion of predators by starling flocks is associated with the triggering and propagation of waves through the flock (5). Even in the absence of predators, we can see deviations of individual behavior from the average behavior of the flock, and correlations in these fluctuations provide a signature of information flow through the flock. Strikingly, observations on flocks of starlings show that these correlations extend over very long distances, comparable to the size of the flock itself (6).

It is generally believed that the interactions among birds in a flock are local—each bird aligns its flight direction and speed to those of its near neighbors (7). If this is correct, then we have to understand how local interactions can generate correlations over much longer distances. In physics, we have two very different mechanisms for local interactions to produce long–ranged correlations. If the system spontaneously breaks a continuous symmetry, for example when all of the spins in a magnet select a particular direction in space along which the macroscopic magnetization will point, then the fluctuations in the system are dominated by Goldstone modes that do not decay on any fixed length scale (8). If we can think of the alignment of flight directions in a flock as being like the alignment of spins in a magnet (9–11), then we can understand the emergence of scale-free correlations via Goldstone's theorem. We have shown that this is more than a metaphor (12): the minimally structured model consistent with the observed correlations among flight directions of neighboring birds is equivalent to a model of spins in a magnet, and the resulting (parameter-free) prediction of long-ranged correlations among fluctuations in flight direction agrees quantitatively with the data.

Not just the fluctuations in flight direction, but also the fluctuations in flight speed are correlated over long distances (6). Now there are no Goldstone modes, because choosing a speed does not correspond to breaking any plausible symmetry of the system. However, there is a second mechanism by which physical systems generate scale-free correlations, and this is by tuning parameters to a critical point (8, 13). As we explore the parameter space of a system (e.g., changing temperature and pressure), we encounter phase transitions, where small changes in parameters produce qualitative changes in behavior of a macroscopic sample (e.g., between liquid and gas). Along the lines in parameter space where these phase transitions exist, there are special points, called "critical points," where the dependence on parameters becomes, for very large systems, singular but not discontinuous. At these points, fluctuations (e.g., in the density of the liquid) become correlated on all length scales, from the molecular scale of the interactions to the macroscopic scale of the sample as a whole.

Tuning to a critical point provides a potential explanation for scale-free correlations in speed of flocking birds, but this is just an analogy; the goal of this paper is to construct a quantitative theory. Our strategy follows ref. 12: we construct the least

Significance

The coherent flight of bird flocks is one of nature's most impressive aerial displays. Beyond the fact that thousands of birds fly, on average, with the same velocity, quantitative observations show that small deviations of individual birds from this average are correlated across the entire flock. By learning minimally structured models from field data, we show that these long-ranged correlations are consistent with local interactions among neighboring birds, but only because the parameters of the flock are tuned to special values, mathematically equivalent to a critical point in statistical mechanics. Being in this critical regime allows information to propagate almost without loss throughout the flock, while keeping the variance of individual velocities small.

structured models that are consistent with measured correlations among neighboring birds, and then see if these models can correctly predict the persistence of correlations over much longer distances, comparable to the size of the flock. We will see that this works, and that the underlying mechanism really is the tuning of the system to a critical point. From a biological point of view, this means that individuals in a flock combine individual speed control and social interactions with their neighbors to achieve a maximal range of influence while keeping speed variability low.

Building a Model from Data

We consider flocks of European starlings, *Sturnus vulgaris*, in the field. The work of refs. 14 and 15 provides a detailed description of these flocks, resulting in the assignment of 3D positions and velocities, at each moment in time, to each individual bird in flocks with up to several thousand members (for a summary, see *SI Text*, section I). From these raw data, one can extract a variety of features that serve to characterize the nature of the ordering in the flock (6, 16).

The positions and velocities of all of the birds in the flock are stochastic—with elements of randomness, but correlated. In making a model, we want to predict the probability distribution out of which these random variables are drawn. One approach is to consider a detailed model for the dynamics of the flock, typically with many parameters to describe the interactions that cause the flock to cohere and align. However, the connection between the model dynamics and the joint distribution of velocities in the flock can be complicated, and fitting the parameters of the interactions is difficult (17); there also is a problem of whether we should take such models seriously as description of the "microscopic" interactions among individuals, or whether they are to be considered as effective interactions in the spirit of statistical physics. As an alternative, we can take some set of observations on the flock as given and try to construct models that reproduce these observations exactly; among the (generally infinite) set of models that can do this, we want to choose the one that has the least structure. Minimizing structure means that the velocities we choose out of the distribution are as random as they can be while still matching the properties of the flock that we have chosen as essential. As emphasized by Jaynes (18, 19), these minimally structured distributions have maximum entropy, providing a connection to the ideas of statistical physics (*SI Text*, section II).

A realistic model for a flock might or might not lead to a maximum entropy distribution. Also, the maximum entropy method is not in itself a model: to use the method we have to choose some set of experimental observations as constraints, and it is easy to imagine choosing the wrong ones. Thus, the maximum entropy approach is a source of hypotheses, and these must be tested. If the maximum entropy model consistent with a limited set of experimental constraints is accurate, correctly predicting experimental observations that were not part of its formulation, then we can take the model seriously and ask what it teaches us about the system.

The maximum entropy approach to model building is far from new, but there has been a resurgence of interest in the use of these ideas to describe biological systems (20–29). In ref. 12, we took a first step toward a maximum entropy description of flocks, building models for the distribution of flight directions that match the average local correlation between the direction of a bird and its nearest neighbors. Surprisingly, fixing this one number leads to a model that, with no free parameters, provides an essentially complete, quantitative description of the propagation of directional order throughout the entire flock. Here we generalize this approach to consider not just flight directions, but also speed. As explained above, we expect that accounting for the properties of speed ordering is a qualitatively different problem from the case of directional ordering.

Given the positions of the birds in space, the state of the flock is defined by the velocity $\vec{v}_i$ of each bird. This 3D vector is composed of the speed, $v_i \equiv |\vec{v}_i|$, and a unit vector, $\vec{s}_i = \vec{v}_i/v_i$, that points in the direction of flight. Our intuition is that the most important interactions are local, between a bird and its immediate neighbors. If this is correct, then the essential features of the system should be captured by measuring local correlations, as in ref. 12.

We can quantify local correlations in the flock by asking how similar, on average, the velocity of each bird is to its neighbors. To do this, we define

$$Q_{\text{int}} = \frac{1}{2v_0^2 N} \sum_{i=1}^{N} \frac{1}{n_c} \sum_{j \in \mathcal{N}_i} \left| \vec{v}_i - \vec{v}_j \right|^2. \qquad [1]$$

Here $\mathcal{N}_i$ is the relevant neighborhood of bird i, which we take to be its first n_c nearest neighbors (12, 16). We compare a bird to each of its neighbors, average over the neighborhood, and then average over all N birds in the flock; we normalize the result by a typical speed v_0 so that we have a dimensionless measure of correlation or similarity. If we take v_0 to be the average speed of birds in the flock, then typical values for Q_{int} are $\sim 10^{-2}$ (Table S1), showing that birds indeed fly with velocities very similar to those of their neighbors.

The definition of Q_{int} quantifies the similarity of each bird's flight vector to that of its neighbors, but if we add a constant to all of the velocities, so that the flock flies faster or slower, then Q_{int} is unchanged. We would like to fix the average speed of the birds in the flock, $V = (1/N)\sum_{i=1}^{N} v_i$, to its observed value $\langle V \rangle_{\text{exp}}$. In addition, we know that individual birds have speeds that vary around the mean, so we would also like to match the variance of speeds. This is equivalent to fixing the mean square speed, $V_2 = (1/N)\sum_{i=1}^{N} v_i^2$. In what follows, we will refer also to the fractional variance in speed,

$$\sigma^2 = \frac{1}{NV^2} \sum_{i=1}^{N} (v_i - V)^2. \qquad [2]$$

The maximum entropy distribution consistent with measured values of Q_{int}, V, and V_2 has the form (*SI Text*, section III),

$$P(\{\vec{v}_i\}) = \frac{1}{Z} \exp\left[-\frac{J}{4v_0^2} \sum_{ij=1}^{N} n_{ij} \left| \vec{v}_i - \vec{v}_j \right|^2 + \frac{\mu}{v_0} \sum_{i=1}^{N} v_i - \frac{g}{2v_0^2} \sum_{i=1}^{N} v_i^2 \right],$$
$$[3]$$

where Z is a constant that ensures the normalization of the probability distribution, and we have inserted factors of v_0 so that other parameters are dimensionless. The matrix n_{ij} maps the connections between birds: $\hat{n}_{ij} = 1$ if bird j is in the neighborhood of bird i ($j \in \mathcal{N}_i$), and zero otherwise; we symmetrize to give $n_{ij} = (\hat{n}_{ij} + \hat{n}_{ji})/2$. The parameters J, μ, and g must be adjusted so that the average values of Q_{int}, V, and V_2 computed from the probability distribution match those observed for the flock; as explained in *SI Text*, section IV, these computations can be done analytically. The only remaining parameter is the number of relevant neighbors n_c, which we fix by requiring that the probability of the observed velocities be as large as possible. We expect that birds on the boundary of the flock will experience different signals than those in the interior; rather than making an explicit model, we fix the velocities of the boundary birds (12), so that Eq. 3 provides a theory for the propagation of order through the flock.

Fig. 1 shows one example of our solution to the inverse problem of determining the parameters J, g, and n_c. Importantly, the quantities that we are trying to match are averages over all of the birds in the flock, and so they are determined with small errors even from a single snapshot of the velocities. The parameters in turn are
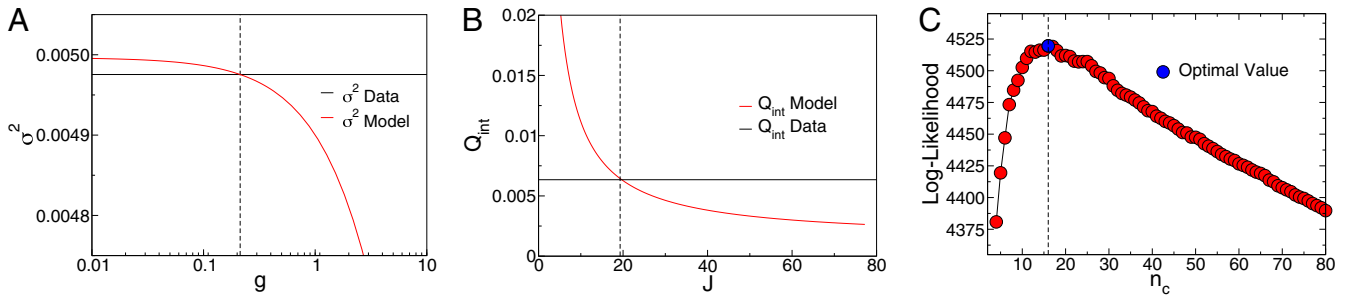
Fig. 1. Inference of the three interaction parameters g, J, and n_c. (A) For fixed values of J and n_c, the value of the speed control parameter g is found by equating the theoretical prediction for the variance of fractional speed fluctuations, σ^2 (red line) from Eq. **2**, to its experimental value (black horizontal line). (B) Once the value of g is determined for all possible values of J and n_c, the interaction strength J can be set by equating the theoretical prediction for Q_{int} (red line) to its experimental value (black horizontal line). (C) Once g and J are computed for given values of n_c, the log-likelihood of the data, $\langle \ln P(\{\vec{v}_i\}) \rangle_{exp}$ becomes a function of n_c only, and the interaction range n_c can be evaluated by maximizing this function. All panels refer to the same single snapshot of one flock (frame 2 of 25-10 in Tables S1 and S2), and mathematical details can be found in *SI Text*, section IV.

determined very precisely, and are consistent for a single flock across time and across multiple flocking events, as in ref. 12 and *SI Text*, section V. In particular, values of n_c are independent of the size of the flock (Fig. S1) and of the distance between birds (Fig. S2), supporting the idea that birds interact with a fixed number of neighbors.

Some Intuition

Maximum entropy distributions are mathematically equivalent to the Boltzmann distribution for systems in thermal equilibrium, and we can use this identity to gain some intuition for the predictions of the model. We recall that a system described by the Boltzmann distribution will occupy a state s with probability $P_s \propto \exp(-E_s/k_BT)$, where E_s is the energy of the state and k_BT is the typical thermal energy; for our purposes we can choose units so that $k_BT = 1$. Thus, Eq. **3** defines an energy function or Hamiltonian on the space of the birds' velocities, and this can be written as

$$\mathcal{H}(\{\vec{v}_i\}) = \frac{J}{4V^2}\sum_{ij=1}^{N} n_{ij}|\vec{v}_i - \vec{v}_j|^2 + \frac{g}{2V^2}\sum_{i=1}^{N}(v_i - V)^2, \quad \textbf{[4]}$$

where we have eliminated the parameter μ in favor of the mean speed V, which is now fixed to its experimental value $\langle V \rangle_{exp}$, and we have set the arbitrary scale $v_0 = V$.

The first term in this Hamiltonian describes the tendency of the individual velocities to adjust both direction and modulus to their neighbors, while the second term forces the speed to have, on average, the value V. From this perspective, we can interpret J as the stiffness of an effective "spring" that ties each bird's velocity to that of its neighbors, and g as the stiffness of a competing spring that ties each speed to the desired mean. Larger J means a tighter connection to the neighbors, and larger g means a tighter individual control over speed.

There are interesting limiting cases that give us a sense for what this model predicts. If the parameter g is very large, then the speed of individual birds hardly fluctuates at all. In this limit, we can rewrite the Hamiltonian as

$$\mathcal{H}(\{\vec{v}_i\}) \approx \mathcal{H}_{dir}(\{\vec{s}_i\}) = -\frac{J}{2}\sum_{ij=1}^{N} n_{ij}\vec{s}_i \cdot \vec{s}_j. \quad \textbf{[5]}$$

This describes the tendency of individual birds to align with their neighbors, and is exactly the model in ref. 12.

If there are nonzero but small fluctuations in speed, then we can write $v_i = V(1 + \epsilon_i)$, and expand in powers of ϵ. The result (*SI Text*, section IV) is that

$$\mathcal{H}(\{\vec{v}_i\}) \approx \mathcal{H}_{dir}(\{\vec{s}_i\}) + \mathcal{H}_{sp}(\{\epsilon_i\}), \quad \textbf{[6]}$$

where the speed Hamiltonian

$$\mathcal{H}_{sp}(\{\epsilon_i\}) = \frac{1}{2}\sum_{i,j=1}^{N}(g\delta_{ij} + JN_{ij})\epsilon_i\epsilon_j, \quad \textbf{[7]}$$

$$N_{ij} = -n_{ij} + \delta_{ij}\sum_{k=1}^{N} n_{ik}. \quad \textbf{[8]}$$

Thus, our full model breaks into two pieces, one describing fluctuations in flight direction, and one describing fluctuations in speed. However, the strength of the springs that tie the speed of each bird to that of its neighbors is determined by the same parameter J which enters the description of directional fluctuations in Eq. **5**. Thus, we have a unified model for how birds adjust their vector velocities to those of their neighbors, rather than separate models (with separate parameters) for the adjustment of direction and speed.

To get a sense for the structure of $\mathcal{H}_{sp}$, it is useful to imagine a continuum limit in which the variations in speed from bird to bird are so smooth that we can picture the speed fluctuations as a continuous function of position in the flock, $\epsilon(\vec{x})$. In this limit (*SI Text*, section VI), we have

$$\mathcal{H}_{sp} \approx \frac{\rho}{2}\int d^3x \left[Jn_cr_c^2(\nabla\epsilon)^2 + g\epsilon^2(\vec{x}) \right], \quad \textbf{[9]}$$

where r_c is the typical distance to a neighboring bird, and ρ is the density of the flock. This model predicts that

$$\langle \epsilon(\vec{x})\epsilon(\vec{x}') \rangle \propto \exp(-|\vec{x} - \vec{x}'|/\xi_{bulk}), \quad \textbf{[10]}$$

where correlation length

$$\xi_{bulk} \sim r_c\sqrt{Jn_c/g} \quad \textbf{[11]}$$

determines the distance over which the fluctuations in speed will be correlated; the subscript reminds us that we are treating the flock as a bulk material, with no boundaries. In this simple picture, there is a critical point at $g = 0$ where the correlation length ξ_{bulk} becomes infinite.

As we have written our model, we need to have $g > 0$. On the other side of the critical point at $g = 0$, we need to constrain the speed distribution more fully (that is, more than just fixing the mean

and the variance) to have a well-normalized distribution of velocities. If we make these extensions, then $g < 0$ will describe a flock in which there is a bimodal distribution of speeds, which seems unnatural. Thus, in this case, the critical point likely is also the boundary of the biologically relevant parameter space.

To summarize, J determines the propagation of directional order through the flock, and to describe the speed fluctuations we have only one extra parameter g. The value of g is set by matching the observed variance in speed across the birds in the flock (Fig. 1A). However, J and g also compete to determine the distance over which speed fluctuations will be correlated, Eq. **11**. Importantly, we are not free to adjust this correlation length by fitting: either the model gets it right, or it does not.

Scale-Free Correlations

Once the parameters J, g, and n_c are determined (Fig. 1), Eq. **3** provides a model for the joint distribution of velocities for all of the birds in the flock; everything that we compute from this distribution is a parameter-free prediction. We start by measuring the similarity of the vector velocities among birds that are not just nearest neighbors, but are separated by greater distances. By analogy with Eq. **1**, we can define

$$Q(r) = \frac{1}{V^2}\left\langle |\vec{v}_i - \vec{v}_j|^2 \right\rangle_{r_{ij}=r}, \qquad [12]$$

where the average is over all pairs of birds separated by a distance $r_{ij} = r$. The predicted $Q(r)$ matches the data very closely (Fig. 2A), out to distances comparable to the size of the flock, more than 10 times farther than the nearest neighbors.

We next decompose the relationships among velocities into contributions from direction and speed. If we average all of the unit vectors $\vec{s}_i$ we obtain the overall polarization of the flock,

$$\vec{P} = \frac{1}{N}\sum_{i=1}^{N}\vec{s}_i, \qquad [13]$$

and we can characterize the fluctuations around this overall direction by a correlation function

$$C_{dir}(r) = \left\langle (\vec{s}_i - \vec{P}) \cdot (\vec{s}_j - \vec{P}) \right\rangle_{r_{ij}=r}. \qquad [14]$$

In Fig. 2B we compare the data with the predictions of the model, and again find good agreement on all scales.

By analogy with Eq. **14**, we can define correlations among the fluctuations in speed,

$$C_{sp}(r) = \left\langle (v_i - V) \cdot (v_j - V) \right\rangle_{r_{ij}=r}. \qquad [15]$$

Fig. 2C shows that the observed correlations are in agreement with the predictions of the model, again over the full range of distances. Thus, we have succeeded in constructing a model based on local interactions that generates correlations over long distances, matching the data quantitatively.

The discussion above suggests that long-ranged correlations are associated with the approach to a critical point at $g = 0$. To see if this intuition is correct, we show in Fig. 3A what happens to the predicted $C_{sp}(r)$ as we change g. Large values of g correspond to small variances in speed, and to correlation functions that decay very rapidly with distance. As g becomes smaller, both the speed variance and the correlation length increase, until, for sufficiently small g, there really is no characteristic scale to the decay of the correlations, and $C_{sp}(r)$ is almost a straight line. This is what we observe, and the success of the theory is that the value of g that matches the observed speed variance is in this regime.

We can quantify the approach to criticality by the dimensionless ratio $g/(Jn_c)$ that enters Eq. **11**. From Fig. 1, we see that $g/(Jn_c) \sim 10^{-3}$, and this is typical (Table S2). This suggests that real flocks are very close to criticality, and that this is why we observe scale-free speed correlations. Note that g cannot be exactly zero, otherwise there is nothing to fix the mean speed of the flock (Eq. **3**), and hence the variance in speed relative to a fixed observer (i.e., ground speed) would be infinite. In contrast, the model predicts that the variance of individuals relative to the flock remains finite as $g \to 0$, and the actual value is quite small, $\sigma^2 \sim 0.005$ in Fig. 1. This measured value of σ^2, together with Q_{int}, fixes the ratio $g/(Jn_c)$ to be small enough to generate scale-free correlations.

To be more precise we need to take into account the finite size of the flocks. Eqs. **10** and **11** hold only for an infinite system; for a finite system, the range of the correlation is limited by the system size. As g is lowered, the behavior of the correlations is influenced more and more by these finite size effects: the exponential decay in Eq. **10** is modified, and the typical distance over which correlations extend is no longer described by ξ_{bulk}. A more faithful estimate of the correlation length ξ is given instead by the zero of the correlation function (6), and the theoretical prediction depends in a nontrivial way on g and the system size L. For small enough values of g, however, the system is effectively critical, and we should see $\xi \propto L$. In Fig. 3A we show that decreasing g below the level required to match the speed variance of the real flock has essentially no effect, and curves with smaller values of g "pile up" as shown in yellow. Repeating the analysis on flocks of different sizes (Fig. 3B), the correlation length does scale with size, and this pattern is captured perfectly by our maximum entropy models.
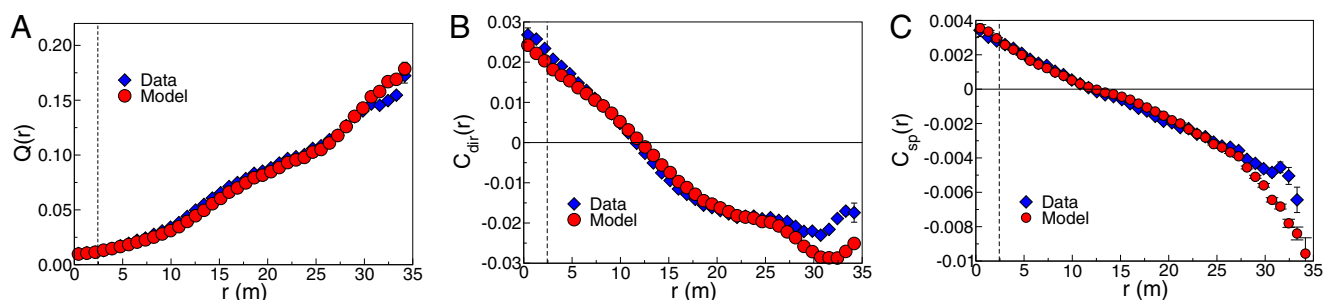
Fig. 2. Correlation functions predicted by the maximum entropy model (red circles) vs. experiments (blue diamonds). (*A*) Similarity of velocities as a function of distance, defined in Eq. **12**. The dashed vertical line indicates the size of the neighborhood defined by n_c birds, within which we match the average Q exactly, by construction. (*B*) Correlations between fluctuations in flight direction as a function of distance, defined in Eq. **14**. (*C*) Correlations between fluctuations in speed as a function of distance, defined in Eq. **15**. All panels refer to the same flock and snapshot as in Fig. 1; theoretical predictions are from Eqs. S88–S91.
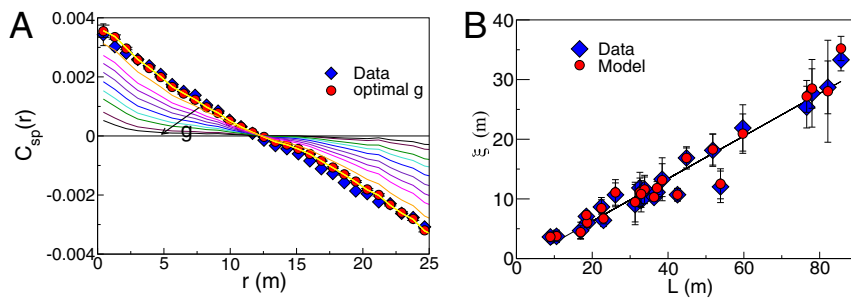
Fig. 3. (*A*) Correlation function of the speed fluctuations for different values of the control parameter g (increasing in the direction of the arrow). All curves with g less than the optimal value collapse onto the yellow curve. (*B*) Correlation length, defined as the point where the correlation function crosses zero (6), in flocks of different sizes, for the experimental data (blue diamonds) and for the model (red circles).

We conclude that flocks exhibit critical behavior, being close enough to the critical point to achieve maximum speed correlation length while maintaining a well-defined cruising speed. These conclusions also hold in more general maximum entropy models where speed and flight directions are regulated by different interaction parameters (*SI Text*, section VII, and Fig. S3).

Dynamical Model

The fact that maximum entropy models are equivalent to the Boltzmann distribution suggests a natural dynamical model, in which the various degrees of freedom in the system execute Brownian motion on the energy landscape:

$$\gamma \frac{d\vec{v}_i(t)}{dt} = -\nabla_i \mathcal{H}\left(\{\vec{v}_j\}\right) + \vec{\eta}_i(t), \qquad [16]$$

where ∇_i indicates the derivatives with respect to the components of the velocity $\vec{v}_i$, γ is a constant to set the time scale of the dynamics, and the Langevin force $\vec{\eta}_i(t)$ is a random, white noise function of time. These dynamics are guaranteed, if the positions of the birds are fixed, to generate velocities that are drawn from the probability distribution in Eq. **3**. However, to give a more realistic model we should add to Eq. **16** forces that depend on the positions of the birds (30–32), so as to fix the overall density of the flock (*SI Text*, section VIII), and the velocities should drive the birds' positions,

$$\frac{d\vec{x}_i}{dt} = \vec{v}_i. \qquad [17]$$

Eqs. **16** and **17** define a self-propelled particle model of interacting birds, and is similar to the Vicsek model, so often used to describe flocking particles (33, 34). In contrast to that model and to most of flocking models in the literature, the speed of the individual particles is not fixed, but regulated by the control parameter g.

Simulations of the dynamical model defined by Eqs. **16** and **17** are shown in Fig. 4. As expected from the (static) maximum entropy model, fluctuations in speed have a correlation length that grows as g is reduced. If g is not too small, correlations decay

exponentially (Eq. **10**), and the correlation length varies with $g/(Jn_c)$ as expected. When g is sufficiently small, the exponential decay is modified by finite size corrections, and the correlation length—now computed as the zero-crossing point of the correlation function—keeps decreasing until a maximal, size-dependent saturation value is reached. In this regime, the correlations extend over a distance determined by the system size, and ξ grows linearly with L, corresponding to scale-free behavior (Fig. 4*B*, *Inset*). This scenario confirms that the mechanism identified in the previous section produces scale-free correlations in the speed even when the full dynamical behavior of the flock is taken into account.

Conclusions

The understanding of collective behavior in matter at thermal equilibrium provides a touchstone for thinking about emergent phenomena in biological systems. Flocking is an especially attractive example, in which the alignment of birds in a flock reminds us of the alignment of spins in a magnet or molecules in a liquid crystal. However, birds are vastly more complex than spins, and this might be nothing more than a metaphor. The goal of this paper and our previous work (12) has been to show that we can go beyond metaphor, that there is a statistical mechanics description of flocks which makes quantitative, parameter-free predictions that are in detailed agreement with the data.

One dramatic collective phenomenon that can emerge in statistical mechanics is a critical point. At such points, distant elements of a system become correlated with one another, far beyond the range of local interactions among the individual elements. At generic parameter values, correlations are expected to decay on some characteristic spatial scale ξ, so that a very large system is composed of many nearly independent pieces of volume ξ^3; often, ξ is not much larger than the range of the interactions themselves. However, at a critical point, the correlation length ξ becomes (formally) infinitely large, and the scale over which correlations extend becomes comparable to the linear size L of the entire system; rather than having many independent pieces, the system acts (almost) as one.

The idea that biological systems might be poised near a critical point is not new (35), but has languished for lack of detailed comparison with experiment. The emergence of more extensive
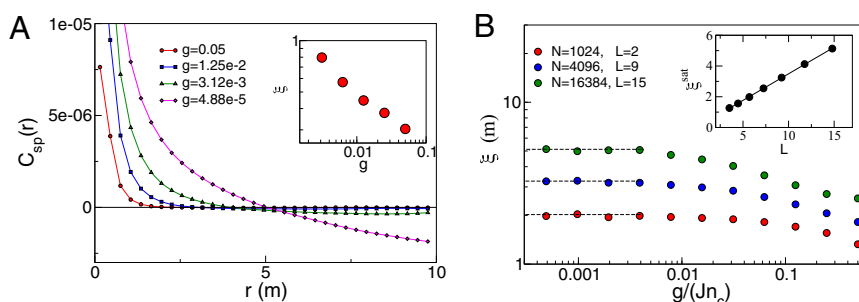

Fig. 4. Simulations of a dynamical model (see *SI Text*, section VIII for details). (*A*) Correlation function of the speed fluctuations at different values of g in a flock of $N = 16,384$ birds. (*Inset*) Correlation length, measured from the exponential decay of the correlation functions at small r, as a function of $g/(Jn_c)$. (*B*) For smaller g, correlation lengths are measured from the zero crossing of the correlation function. For $g/(Jn_c) \ll 1$, ξ approaches a maximum value that depends on the size of the system. (*Inset*) Low g maximum of ξ, as a function of the system size; the linear dependence of ξ on L is typical of scale-free behavior.

data, as well as ideas about how to connect theory and experiment, has led to a reexamination of criticality in a wide variety of biological systems (36). In this context, the observation of long-ranged or scale-free correlations in the velocities of starlings in a flock (6) is very suggestive. Our results here show that these correlations are not just analogous to the correlations at a critical point: we have a very accurate description of the entire distribution of speed and direction fluctuations in the flock, this description is mathematically equivalent to a statistical mechanics model of a magnet, and the observed scale-free correlations are predicted correctly because the parameters of this model are in the critical regime.

Our approach is not a fit to the observed scale-free behavior of the flock. Instead we take from the data a measurement of local correlations, and the variance of individual birds' speeds relative to the average over the flock, and build the least structured model that is consistent with these two measurements. Thus, rather than thinking of criticality as occurring in the neighborhood of a special point in the space of model parameters, we can think of it as a statement about the behavior of the flock itself. In particular, as emphasized in Fig. 3, even a factor-of-2 change in the variance of the speeds would predict correlations that decay much more rapidly with distance, inconsistent with what we see in real flocks.

Biologically, birds may vary their speeds either for individual reasons (37), or to follow their neighbors. In this language, the critical point is the place where social forces overwhelm individual preferences. More broadly, the critical regime is one in which individuals achieve maximal coherence with their neighbors while still keeping some control over their speeds.

Why do flocks organize themselves to be critical? There has been much more speculation about the advantages of criticality for biological systems than there has been direct evidence, so we do not want to add too much here. We note, however, that in the statistical mechanics framework, long-ranged correlations at criticality are mathematically equivalent to the statement that information can propagate over similarly long distances. Away from criticality, a signal visible to one bird on the border of the flock can influence just a handful of near neighbors; at criticality, the same signal can spread to influence the behavior of the entire flock. Such susceptibility seems advantageous, but it would be attractive to have more direct measurements of the propagating signal (1). The critical point is a place where many quantities are extremal; it remains to be seen which of these is most meaningful to the birds.

1. Attanasi, et al. (2013) Superfluid transport of information in turning flocks of starlings. arXiv:1303.7097 [cond-mat.stat-mech].
2. Pulliam HR (1973) On the advantages of flocking. *J Theor Biol* 38(2):419–422.
3. Cresswell W (1994) Flocking is an effective anti-predation strategy in redshanks, Tringa totanus. *Anim Behav* 47(2):433–442.
4. Krause J, Ruxton GD (2002) *Living in Groups* (Oxford Univ Press, Oxford).
5. Procaccini A, et al. (2011) Propagating waves in starling, Sturnus vulgaris, flocks under predation. *Anim Behav* 82(4):759–765.
6. Cavagna A, et al. (2010) Scale-free correlations in starling flocks. *Proc Natl Acad Sci USA* 107(26):11865–11870.
7. Couzin ID, Krause J (2003) Self-organization and collective behavior in vertebrates. *Adv Stud Behav* 32:1–75.
8. Parisi G (1988) *Statistical Field Theory* (Addison-Wesley, Redwood City, CA).
9. 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.
10. 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.
11. Ramaswamy S (2010) The mechanics and statistics of active matter. *Annu Rev Cond Matt Phys* 1:323–345.
12. Bialek W, et al. (2012) Statistical mechanics for natural flocks of birds. *Proc Natl Acad Sci USA* 109(13):4786–4791.
13. Wilson KG (1979) Problems in physics with many scales of length. *Sci Am* 241(2):158–179.
14. Cavagna A, et al. (2008) The STARFLAG handbook on collective animal behaviour: 1. Empirical methods. *Anim Behav* 76(1):217–236.
15. 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(1):237–248.
16. 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.
17. Gautrais J, et al. (2012) Deciphering interactions in moving animal groups. *PLOS Comput Biol* 8(9):e1002678.
18. Jaynes ET (1957) Information theory and statistical mechanics. *Phys Rev* 106:620–630.
19. Mackay DJC (2003) *Information Theory, Inference, and Learning Algorithms* (Cambridge Univ Press, Cambridge, UK).
20. Schneidman E, Berry MJ, 2nd, Segev R, Bialek W (2006) Weak pairwise correlations imply strongly correlated network states in a neural population. *Nature* 440(7087):1007–1012.
21. Shlens J, et al. (2006) The structure of multi-neuron firing patterns in primate retina. *J Neurosci* 26(32):8254–8266.
22. Lezon TR, Banavar JR, Cieplak M, Maritan A, Fedoroff NV (2006) Using the principle of entropy maximization to infer genetic interaction networks from gene expression patterns. *Proc Natl Acad Sci USA* 103(50):19033–19038.
23. Tang A, et al. (2008) A maximum entropy model applied to spatial and temporal correlations from cortical networks in vitro. *J Neurosci* 28(2):505–518.
24. Weigt M, White RA, Szurmant H, Hoch JA, Hwa T (2009) Identification of direct residue contacts in protein-protein interaction by message passing. *Proc Natl Acad Sci USA* 106(1):67–72.
25. Halabi N, Rivoire O, Leibler S, Ranganathan R (2009) Protein sectors: Evolutionary units of three-dimensional structure. *Cell* 138(4):774–786.
26. Mora T, Walczak AM, Bialek W, Callan CG, Jr. (2010) Maximum entropy models for antibody diversity. *Proc Natl Acad Sci USA* 107(12):5405–5410.
27. Stephens GJ, Bialek W (2010) Statistical mechanics of letters in words. *Phys Rev E Stat Nonlin Soft Matter Phys* 81(6 Pt 2):066119.
28. Tkačik G, Marre O, Mora T, Amodei D, Berry MJ II, Bialek W (2013) The simplest maximum entropy model for collective behavior in a neural network. arXiv:1207.6319 [q-bioNC].
29. Tkačik G, et al. (2013) Searching for collective behavior in a network of sensory neurons. *PLoS Comput Biol* 10(1):e1003408.
30. Grégoire G, Chaté H (2004) Onset of collective and cohesive motion. *Phys Rev Lett* 92(2):025702.
31. 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.
32. Pohl O (2011) *Analyse und Simulation eines stochastischen Modells zur Schwarmdynamik* (Diplomarbeit, Rheinischen Friedrich-Wilhelms-Universität Bonn, Bonn).
33. 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.
34. Grégoire G, Chaté H, Tu Y (2003) Moving and staying together without a leader. *Physica D* 181(3-4):157–170.
35. Bak P (1996) *How Nature Works: The Science of Self-Organized Criticality* (Copernicus, New York).
36. Bialek W, Mora T (2011) Are biological systems poised at criticality? *J Stat Phys* 144(2):268–302.
37. Rayner JMV, Viscardi PW, Ward S, Speakman JR (2001) Aerodynamics and energetics of intermittent flight in birds. *Am Zool* 41(2):188–204.

PHYSICS

BIOPHYSICS AND
COMPUTATIONAL BIOLOGY

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, $\{\langle O_\mu(\mathbf{v})\rangle_{\exp}\}$.

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

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

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 $\{\langle O_\mu(\mathbf{v})\rangle_{\exp}\}$. 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),

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

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,

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

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

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

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,

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

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

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

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

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

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

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

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

$$P_{\text{model}}(\text{data}) = \prod_{i=1}^{N_s} P(\mathbf{v}^{(i)}). \qquad \textbf{[S9]}$$

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

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

Then we can form the normalized log probability,

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

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

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:

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

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

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

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

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

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

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

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

and hence Eq. S13 becomes

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

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 $\mathbf{v} \equiv \{\vec{v}_1, \vec{v}_2 \cdots, \vec{v}_N\}$ 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,

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

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

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

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

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

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

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

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

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

where $v_i = |\vec{v}_i|$ 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

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

and the probability distribution

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

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

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

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_{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

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

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

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, $\vec{v}_i = v_i \vec{s}_i$, where $|\vec{s}_i| = 1$. Translating into these variables, we obtain from Eq. **S27**:

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

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

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

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,

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

where the delta function enforces the constraint that $\vec{s}_i$ 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

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

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

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

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

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

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

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

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

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

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

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

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

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

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 $\vec{s}_i$ 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

$$P = \frac{1}{N}\sum_{i=1}^{N} s_i^L, \qquad \text{[S40]}$$

and

$$\sum_{i=1}^{N} \vec{\pi}_i = 0. \qquad \text{[S41]}$$

Thus, we have

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

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

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

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

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

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

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

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,

$$\mathcal{H}(\mathbf{v}) = \mathcal{H}_{\mathrm{dir}}\left(\{\vec{\pi}_i\}\right) + \mathcal{H}_{\mathrm{sp}}(\{\epsilon_i\}) + E_0(V), \qquad \text{[S46]}$$

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

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

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

$$\mathcal{H}_{\mathrm{dir}}\left(\{\vec{\pi}_i\}\right) = \frac{JV^2}{2v_0^2}\sum_{i,j=1}^{N} N_{ij}\vec{\pi}_i \cdot \vec{\pi}_j \qquad \text{[S48]}$$

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

where the matrix N_{ij} has the form

$$N_{ij} = -n_{ij} + \delta_{ij}\sum_{k=1}^{N} n_{ik}. \qquad \text{[S50]}$$

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 $\hat{n}$. The integral over the direction of polarization is simple because there is no dependence of the integrand on $\hat{n}$; 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

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

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

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

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 $V \rightarrow V_* = \langle V \rangle_{\exp}$ everywhere else in our calculations. We are also free to choose the scale $v_0 = \langle V \rangle_{\exp}$, and then we can simplify

$$\mathcal{H}_{\mathrm{dir}}\left(\{\vec{\pi}_i\}\right) = \frac{J}{2}\sum_{i,j=1}^{N} N_{ij}\vec{\pi}_i \cdot \vec{\pi}_j, \qquad \text{[S53]}$$

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

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

$$\mathcal{Z}(J, g, \mu) \propto \mathcal{Z}_{\mathrm{dir}}(J)\mathcal{Z}_{\mathrm{sp}}(J, g)e^{Ng/2}, \qquad \text{[S55]}$$

where

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

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

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_{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_{ij} is the discrete approximation to the (negative) second derivative along the line. In higher dimensions this becomes the Laplacian operator, and so N_{ij} is called a "Laplacian matrix." As with the negative Laplacian, the eigenvalues $\{\Lambda_a\}$ of N_{ij} are positive, except for the smallest one, which is exactly zero ($\Lambda_1 = 0$). If we define the eigenvectors of N_{ij} by w_i^a such that

$$\sum_{j=1}^{N} N_{ij} w_j^a = \Lambda_a w_i^a, \qquad \textbf{[S58]}$$

then the eigenvector associated with the zero eigenvalue is the uniform mode, $w_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_{ij},

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

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

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

The relevant matrix is now $M_{ij} = g\delta_{ij} + JN_{ij}$, and the eigenvalues are $\lambda_a(M) = g + J\Lambda_a$, where again Λ_a are the eigenvalues of the Laplacian matrix N_{ij}. We note that the integral runs over N dimensions, but the delta function fixes one combination of the $\{\epsilon_i\}$ 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

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

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

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

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_{ij}, and note that in this basis fluctuations along each coordinate are independent with variance $1/\lambda_n(M)$; 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_i^1. The end result is that

$$\langle \epsilon_i \epsilon_j \rangle = \sum_{a=2}^{N} \frac{w_i^a w_j^a}{g + J\Lambda_a}. \qquad \textbf{[S63]}$$

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. 2C of the main text.

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

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

and

$$\langle \vec{\pi}_i \cdot \vec{\pi}_j \rangle = (d-1) \sum_{a=2}^{N} \frac{w_i^a w_j^a}{J\Lambda_a}, \qquad \textbf{[S65]}$$

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

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

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

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

where $\langle \cdots \rangle$ denotes an average over the data. We have identified Q_{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

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

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

and similarly

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

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

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

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. 2C 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

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

where

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

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

We can see from these expressions that holding velocities $\vec{v}_a$ fixed on the border of the flock is equivalent to considering a flock in presence of a field $\vec{h}_i$ 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,

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

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

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

and in place of Eqs. S56 and S57 we have

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

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

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 $v_0 = \langle V \rangle_{\text{exp}}$, simplifying all of the expressions. The reduced Hamiltonians for the internal variables, analogs of Eqs. S53 and S54, then become

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

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

where

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

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,

$$\ln \mathcal{Z}_{\text{dir}}(J) = \frac{J}{2} \sum_{i,j\in\mathcal{I}} (A^{-1})_{ij} \vec{h}_i^\perp \cdot \vec{h}_j^\perp - \frac{d-1}{2}(N_\mathcal{I} - 1)\ln(J)$$

$$- \frac{d-1}{2}\ln\left[\sum_{i,j\in\mathcal{I}} (A^{-1})_{ij}\right] - \frac{d-1}{2}\ln \det A$$

$$- \frac{J}{2}\left|\sum_{a\in\mathcal{B}} \vec{\pi}_a + \sum_{i,j\in\mathcal{I}} (A^{-1})_{ij} \vec{h}_i^\perp\right|^2 \frac{1}{\sum_{i,j\in\mathcal{I}} (A^{-1})_{ij}}, \qquad \text{[S84]}$$

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

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

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

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

$$\frac{1}{J}\left(d\frac{N_\mathcal{I}-1}{N} - g\langle\sigma^2\rangle_{exp}\right) = n_c\langle Q_{int}\rangle_{exp} + \frac{1}{N}\sum_{i,j\in\mathcal{I}} (A^{-1})_{ij}\vec{h}_i^\perp \cdot \vec{h}_j^\perp$$

$$+\frac{1}{N}\sum_{i,j\in\mathcal{I}} (B^{-1})_{ij} b_i b_j - \frac{1}{N}\left|\sum_{a\in\mathcal{B}} \vec{\pi}_a + \sum_{i,j\in\mathcal{I}} (A^{-1})_{ij}\vec{h}_j^\perp\right|^2 \frac{1}{\sum_{i,j\in\mathcal{I}} (A^{-1})_{ij}}$$

$$-\frac{1}{N}\left(\sum_{a\in\mathcal{B}} \epsilon_a + \sum_{i,j\in\mathcal{I}} (B^{-1})_{ij} b_j\right)^2 \frac{1}{\sum_{i,j\in\mathcal{I}} (B^{-1})_{ij}}$$

$$-\frac{1}{2N\langle V\rangle_{exp}^2}\sum_{a,b\in\mathcal{B}} n_{ab}|\vec{v}_a - \vec{v}_b|^2 + \frac{g}{NJ\langle V\rangle_{exp}^2}\sum_{a\in\mathcal{B}} \left(v_a - \langle V\rangle_{exp}\right)^2, \quad \textbf{[S92]}$$

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**,

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

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

$$\mathcal{H}_{\mathrm{dir}}\big(\{\vec{\pi}_i\}\big) = \frac{J}{2}\sum_{n=1}^{N}|\vec{\pi}_n - \vec{\pi}_{n+1}|^2. \qquad \textbf{[S95]}$$

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 $x_n = nr_c$, where r_c is the typical distance between the nearest birds. Then we have $\vec{\pi}(x)$, and

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

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

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

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

$$\mathcal{H}_{\mathrm{dir}}\big(\{\vec{\pi}_i\}\big) = \frac{Jn_c r_c^2}{2}\rho\int d^3x|\nabla\vec{\pi}(x)|^2, \qquad \textbf{[S98]}$$

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 $\mathcal{H}_{\mathrm{dir}}\propto A^2/\ell^2$. 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

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

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

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

and then Eq. **S103** becomes

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

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

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

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

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

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

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

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

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

$$\mathcal{H}_{\mathrm{dir}}\big(\{\vec{\pi}_i\}\big) = \frac{J}{2}\sum_{i,j=1}^{N}N_{ij}\vec{\pi}_i\cdot\vec{\pi}_j$$
$$\rightarrow \frac{Jn_c r_c^2}{2}\rho\int d^3x|\nabla\vec{\pi}(x)|^2. \qquad \textbf{[S110]}$$

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

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

$$\rightarrow \frac{1}{2} \rho \int d^3x \left[Jn_c r_c^2 \left| \nabla \epsilon(\vec{x}) \right|^2 + g\epsilon^2(\vec{x}) \right] \qquad \text{[S111]}$$

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

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,

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

Thus, where $S_{\vec{\pi}}$ grows without bound as the wave vector $\vec{k}$ becomes small, $S_\epsilon(\vec{k})$ stops growing once $\vec{k}$ 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

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

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

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

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_{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

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

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

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

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**,

$$\mathcal{H}(\mathbf{v}) = \frac{J^{\text{sp}}}{4v_0^2} \sum_{i,j=1}^{N} n_{ij}^{\text{sp}} (v_i - v_j)^2 - \frac{J^{\text{dir}}}{2v_0^2} \sum_{i,j=1}^{N} n_{ij}^{\text{dir}} \vec{s}_i \cdot \vec{s}_j$$

$$+ \frac{g}{2v_0^2} \sum_{i=1}^{N} v_i^2 - \frac{\mu}{v_0} \sum_{i=1}^{N} v_i, \qquad \text{[S119]}$$

where n_{ij}^{sp} is defined as n_{ij} above, but with neighborhoods of size n_c^{sp}, and similarly for n_{ij}^{dir}. Notice that we now have two different coupling strengths, J^{sp} and J^{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 $J^{\text{sp}} = J^{\text{dir}}$ we are back to our original model. With the two separate parameters we find the log-likelihood, by analogy with Eq. **S94**,

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

$$+ \frac{J^{\text{dir}} n_c^{\text{dir}} N}{2} \langle C_{\text{int}} \rangle_{\text{exp}} - N \frac{g}{2} \langle \sigma^2 \rangle_{\text{exp}}. \qquad \text{[S120]}$$

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. S3A): 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. S3B), 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. S3A) 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

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

$$= -\frac{J}{2v_0^2} \sum_j n_{ij}(\vec{v}_i - \vec{v}_j) - \frac{g}{v_0^2} \frac{\vec{v}_i}{v_i}(v_i - \hat{v})$$

$$+ \frac{1}{n_c} \sum_{j \in \mathcal{N}_i} \vec{f}_{ij} + \vec{\eta}_i(t) \qquad \text{[S122]}$$

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

where we have added, as described in the text, forces $\vec{f}_{ij}$ 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

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

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

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

where $\vec{r}_{ij}$ is the vector from bird i to bird j, $r_{ij} = |\vec{r}_{ij}|$ 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 $\vec{f}_{ij}$, 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.
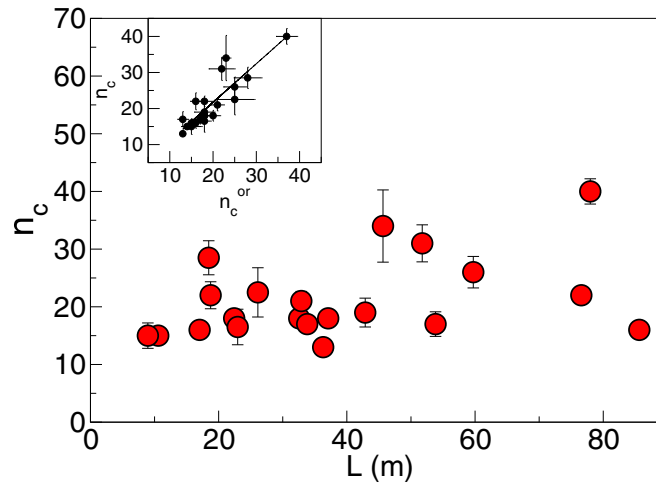
Fig. S1. Inferred values of n_c for flocks of different linear dimension. *Inset* shows that n_c determined here agrees with earlier results for a model of directional ordering alone (7). Error bars indicate SDs across multiple snapshots in each flocking event.
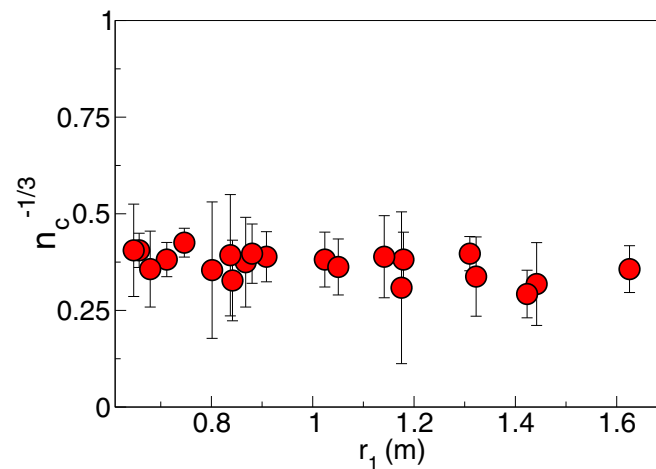

Fig. S2. Inferred values of n_c for flocks as a function of the typical distance r_1 between neighboring birds. Error bars same as in Fig. S1.
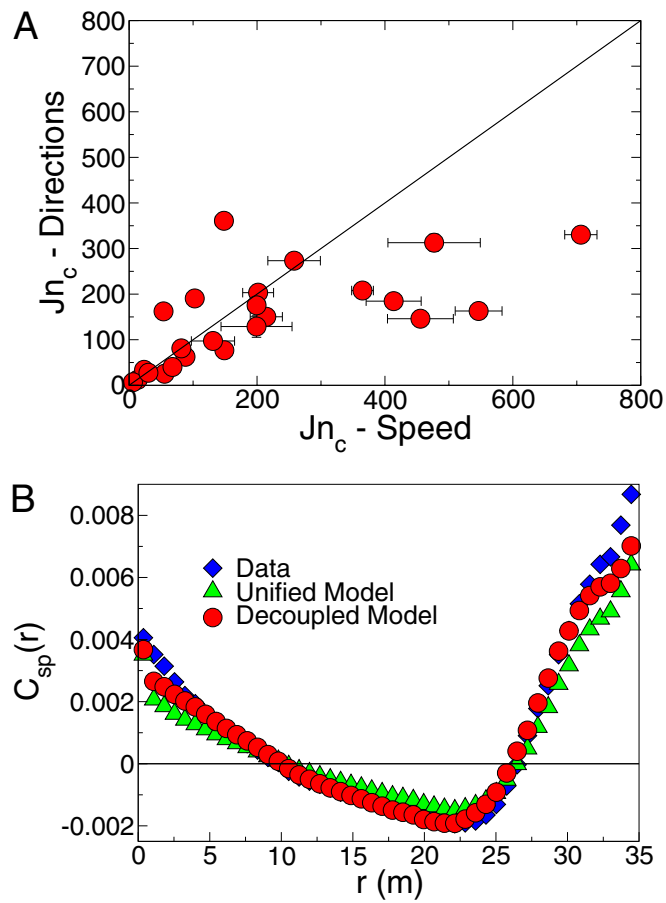
Fig. S3. Model with independent interactions for speed and flight directions. (*A*) The inferred global interaction strength Jn_c for the orientational degrees of freedom (y axis) vs. the speed degrees of freedom (x axis). The straight line corresponds to $y = x$, i.e., to the global model where the 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.