

# Statistical mechanics for natural flocks of birds

William Bialek<sup>a</sup>, Andrea Cavagna<sup>b,c</sup>, Irene Giardina<sup>b,c,1</sup>, Thierry Mora<sup>d</sup>, Edmondo Silvestri<sup>b,c</sup>,  
Massimiliano Viale<sup>b,c</sup>, and Aleksandra M. Walczak<sup>e</sup>

<sup>a</sup>Joseph Henry Laboratories of Physics and Lewis–Sigler Institute for Integrative Genomics, Princeton University, Princeton, NJ 08544; <sup>b</sup>Istituto dei Sistemi Complessi, Consiglio Nazionale delle Ricerche, Rome, Italy; <sup>c</sup>Dipartimento di Fisica, Università Sapienza, Rome, Italy; <sup>d</sup>Laboratoire de Physique Statistique de l'École Normale Supérieure, Centre National de la Recherche Scientifique and University Paris VI, Paris, France; and <sup>e</sup>Laboratoire de Physique Théorique de l'École Normale Supérieure, Centre National de la Recherche Scientifique and University Paris VI, Paris, France

Edited by Boris I. Shraiman, University of California, Santa Barbara, CA, and approved January 28, 2012 (received for review November 18, 2011)

**Flocking is a typical example of emergent collective behavior, where interactions between individuals produce collective patterns on the large scale. Here we show how a quantitative microscopic theory for directional ordering in a flock can be derived directly from field data. We construct the minimally structured (maximum entropy) model consistent with experimental correlations in large flocks of starlings. The maximum entropy model shows that local, pairwise interactions between birds are sufficient to correctly predict the propagation of order throughout entire flocks of starlings, with no free parameters. We also find that the number of interacting neighbors is independent of flock density, confirming that interactions are ruled by topological rather than metric distance. Finally, by comparing flocks of different sizes, the model correctly accounts for the observed scale invariance of long-range correlations among the fluctuations in flight direction.**

animal groups | statistical inference

The collective behavior of large groups of animals is an imposing natural phenomenon, very hard to cast into a systematic theory (1). Physicists have long hoped that such collective behaviors in biological systems could be understood in the same way as we understand collective behavior in physics, where statistical mechanics provides a bridge between microscopic rules and macroscopic phenomena (2, 3). A natural test case for this approach is the emergence of order in a flock of birds: Out of a network of distributed interactions among the individuals, the entire flock spontaneously chooses a unique direction in which to fly (4), much as local interactions among individual spins in a ferromagnet lead to a spontaneous magnetization of the system as a whole (5). Despite detailed development of these ideas (6–9), there still is a gap between theory and experiment. Here we show how to bridge this gap by constructing a maximum entropy model (10) based on field data of large flocks of starlings (11–13). We use this framework to show that the effective interactions among birds are local and that the number of interacting neighbors is independent of flock density, confirming that interactions are ruled by topological rather than metric distance (14). The statistical mechanics models that we derive in this way provide an essentially complete, parameter-free theory for the propagation of directional order throughout the flock.

We consider flocks of European starlings, *Sturnus vulgaris*, as in Fig. 1A. At any given instant of time, following refs. 11–13, we can attach to each bird  $i$  a vector velocity  $\vec{v}_i$  and define the normalized velocity  $\vec{s}_i = \vec{v}_i/|\vec{v}_i|$  (Fig. 1B). On the hypothesis that flocks have statistically stationary states, we can think of all these normalized velocities as being drawn (jointly) from a probability distribution  $P(\{\vec{s}_i\})$ . It is not possible to infer this full distribution directly from experiments, because the space of states specified by  $\{\vec{s}_i\}$  is too large. However, what we can measure from field data is the matrix of correlations between the normalized velocities,  $C_{ij} = \langle \vec{s}_i \cdot \vec{s}_j \rangle$ . There are infinitely many distributions  $P(\{\vec{s}_i\})$  that are consistent with the measured correlations, but out of all these distributions, there is one that has minimal structure: It describes a system that is as random as it can be while still matching the

experimental data. This distribution is the one with maximum entropy (10).

It should be emphasized that the maximum entropy principle is not a “modeling assumption;” rather it is the absence of assumptions. Any other model that accounts for the observed correlations will have more structure and hence (explicitly or implicitly) assumes something about the nature of the interactions in the flock beyond what is required to match the data. Of course the fact that the maximum entropy model is minimally structured does not make it correct. It could be, for example, that individual birds set their flight direction by computing a complicated nonlinear combination of the velocities from multiple neighbors, in which case correlations among pairs of birds would be insufficient to capture the essence of the ordering mechanism. We view the maximum entropy model as a powerful starting point, from which, as we will see, we can generate detailed and testable predictions.

The maximum entropy distribution consistent with the directional correlations  $C_{ij}$  is

$$P(\{\vec{s}_i\}) = \frac{1}{Z(\{J_{ij}\})} \exp \left[ \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N J_{ij} \vec{s}_i \cdot \vec{s}_j \right], \quad [1]$$

where  $Z(\{J_{ij}\})$  is the appropriate normalization factor, or partition function; the derivation follows ref. 10, as explained in the *SI Appendix*. Notice that there is one parameter  $J_{ij}$  corresponding to each measured element  $C_{ij}$  of the correlation matrix. To finish the construction of the model, we have to adjust the values of the  $J_{ij}$  to match the experimentally observed  $C_{ij}$ ,

$$\langle \vec{s}_i \cdot \vec{s}_j \rangle_P = \langle \vec{s}_i \cdot \vec{s}_j \rangle_{\text{exp}}, \quad [2]$$

where the symbol  $\langle \cdot \rangle_P$  indicates an average using distribution  $P$  from Eq. 1, whereas  $\langle \cdot \rangle_{\text{exp}}$  indicates an average over many experiments. This matching condition is equivalent to maximizing the likelihood that the model in Eq. 1 will generate the data from which the correlations were computed.

The probability distribution in Eq. 1 is mathematically identical to a model that is familiar from the physics of magnets—the Heisenberg model (5)—in which a collection of spins  $\vec{s}_i$  interact so that their energy (or Hamiltonian) is  $H(\{\vec{s}_i\}) = -(1/2) \sum_{i,j} J_{ij} \vec{s}_i \cdot \vec{s}_j$ ; Eq. 1 then describes the thermal equilibrium or Boltzmann distribution at a temperature  $k_B T = 1$ . In this context, the constants  $J_{ij}$  are the strength of interaction between

Author contributions: W.B., A.C., I.G., T.M., and A.M.W. designed research; W.B., A.C., I.G., T.M., E.S., M.V., and A.M.W. performed research; A.C., I.G., and M.V. analyzed data; and W.B., A.C., I.G., T.M., and A.M.W. wrote the paper.

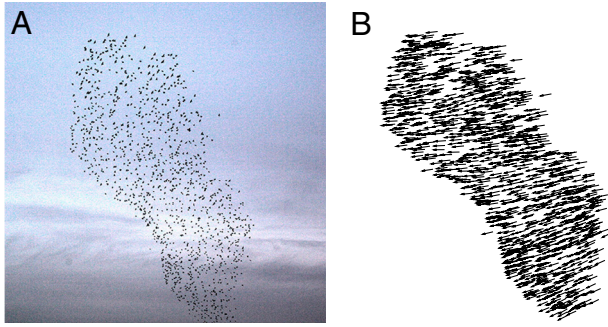
The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

<sup>1</sup>To whom correspondence should be addressed. E-mail: irene.giardina@roma1.infn.it.

This article contains supporting information online at [www.pnas.org/lookup/suppl/doi:10.1073/pnas.1118633109/-DCSupplemental](http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.1118633109/-DCSupplemental).



**Fig. 1.** The raw data. (A) One snapshot from flocking event 28–10,  $N = 1,246$  birds (see [SI Appendix, Table S1](#)). (B) Instantaneous vector velocities of all the individuals in this snapshot, normalized as  $\vec{s}_i = \vec{v}_i/|\vec{v}_i|$ .

spins  $i$  and  $j$ , where  $J > 0$  means that these elements tend to align. For many physical systems, once we know the Hamiltonian there is a plausible dynamics that allows the system to relax toward equilibrium, which is the Langevin dynamics

$$\frac{d\vec{s}_i}{dt} = -\frac{\partial H}{\partial \vec{s}_i} + \vec{\eta}_i(t) = \sum_{j=1}^N J_{ij}\vec{s}_j + \vec{\eta}_i(t), \quad [3]$$

where  $\vec{\eta}_i(t)$  is an independent white noise “force” driving each separate degree of freedom. Finding trajectories  $\vec{s}_i(t)$  that solve Eq. 3 produces samples that are drawn out of the probability distribution in Eq. 1. The interesting point is that this kind of dynamical model also is well known in biology: The direction of motion of an individual evolves in time according to “social forces” reflecting a weighted sum of inputs from neighboring individuals, plus noise (4). In this framework,  $J_{ij}$  measures the strength of the force that tries to align the velocity of bird  $i$  along the direction defined by bird  $j$ . We emphasize that we are doing an analogy but a mathematical equivalence.

In contrast to most networks, the connectivity in a flock of birds is intrinsically dynamic—birds move and change their neighbors. Thus, it may not make sense to talk about matrix of correlations  $C_{ij}$  or interactions  $J_{ij}$  between labeled individuals. On the other hand, the continuous and rapid change of neighbors induced by motion implies that the interaction  $J_{ij}$  between bird  $i$  and bird  $j$  cannot depend directly on their specific identities but only on some function of their relative positions.

The simplest form of interaction that is independent of the birds’ identity is one in which each bird interacts with the same strength,  $J$ , with the same number of neighbors,  $n_c$  (or with all birds within the same radius  $r_c$ ; see below). If the interactions are of this form, then Eq. 1 simplifies to

$$P(\{\vec{s}_i\}) = \frac{1}{Z(J, n_c)} \exp \left[ \frac{J}{2} \sum_{i=1}^N \sum_{j \in n_c^i} \vec{s}_i \cdot \vec{s}_j \right], \quad [4]$$

where  $j \in n_c^i$  means that bird  $j$  belongs to the first  $n_c$  nearest neighbors of  $i$ . It is important to note that Eq. 4 can be also derived, without any assumption, as a maximum entropy distribution consistent with an experimental quantity simpler than the full correlation matrix. Instead of the correlation  $C_{ij}$  for each pair of individuals, we can measure its average value over birds within a neighborhood of size  $n_c$ , i.e.,

$$C_{\text{int}} = \frac{1}{N} \sum_{i=1}^N \frac{1}{n_c} \sum_{j \in n_c^i} \langle \vec{s}_i \cdot \vec{s}_j \rangle \approx \frac{1}{N} \sum_{i=1}^N \frac{1}{n_c} \sum_{j \in n_c^i} \vec{s}_i \cdot \vec{s}_j. \quad [5]$$

It can be easily shown (see [SI Appendix](#)) that the maximum entropy distribution consistent with this scalar correlation is in fact the distribution [4]. As in the more general problem, finding the

values of  $J$  and  $n_c$  that reproduce the observed correlation  $C_{\text{int}}$  is the same as maximizing the probability, or likelihood, that model Eq. 4 generates the observed configuration of flight directions  $\{\vec{s}_i\}$  in a single snapshot. Biologically, Eqs. 4 and 5 encapsulate the concept that the fundamental correlations are between birds and their directly interacting neighbors; all more distant correlations should be derivable from these correlations. If this concept is correct, a model as in Eq. 4 that appropriately reproduces the fundamental correlations up to the scale  $n_c$  must be able to describe correlations on all length scales.

Importantly, with large flocks we can estimate the correlations among interacting neighbors from a single snapshot of the birds’ flight directions  $\{\vec{s}_i\}$ , as indicated in the second step of Eq. 5. In contrast, if we were trying to estimate the entire correlation matrix in Eq. 2, we would need as many samples as we have birds in the flock (see [SI Appendix](#)), and we would have to treat explicitly the dynamic rearrangements of the interaction network during flight. This situation is an extreme version of the general observation that the sampling problems involved in the construction of maximum entropy models can be greatly reduced if we have prior expectations that constrain the structure of the interaction matrix (15, 16).

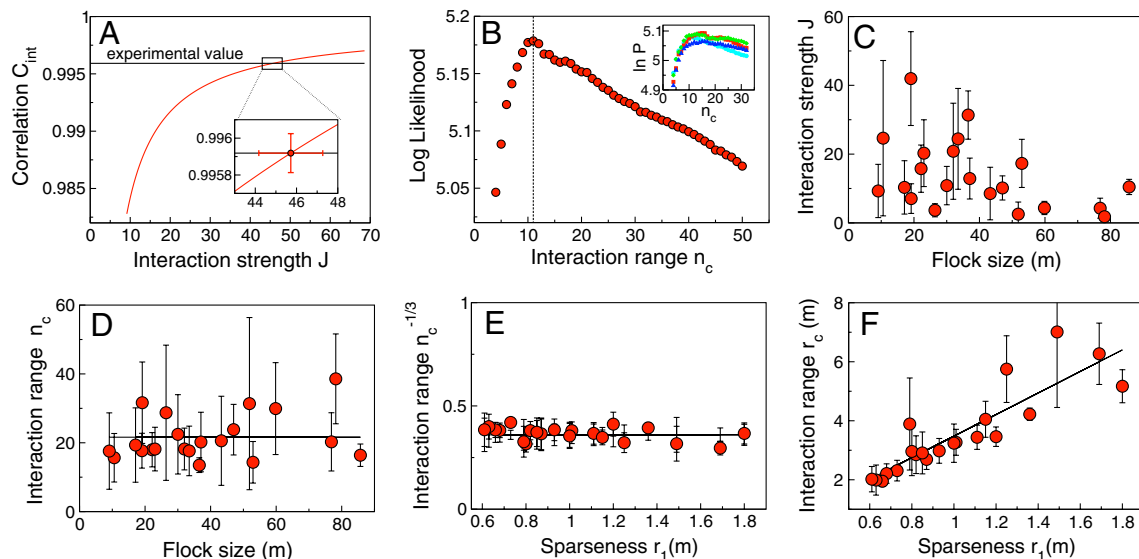
## Results

We now apply this analysis to data on real flocks of starlings. Given a snapshot of the flock, we have the configuration  $\{\vec{s}_i\}$ , and we need to evaluate the probability  $P(\{\vec{s}_i\})$  in Eq. 4 for any value of  $J$  and  $n_c$ , then maximize this probability with respect to these parameters (see [Materials and Methods](#) and [SI Appendix](#) for details of the computation). Special care must be devoted to birds on the outer edge, or border, of the flock, because these individuals have very asymmetric neighborhoods and may receive inputs from signals outside the flock. If we take the flight directions of these border birds as given, we can study how information propagates through the flock, without having to make assumptions about how the boundary is different from the interior. Technically, then, we describe the flock with Eq. 4 but with the flight directions of the border birds fixed (again, see [Materials and Methods](#) and [SI Appendix](#) for details).

**Inferring the Interaction Parameters from Data.** We proceed as follows. For a single flock, at a given instant of time, we compute the correlation  $C_{\text{int}}$  predicted by the model in Eq. 4 as a function of the coupling strength  $J$  and compare it with the experimental value of the correlation (Fig. 2A). The equation  $C_{\text{int}}(J, n_c) = C_{\text{int}}^{\text{exp}}$  fixes  $J(n_c)$  for each value of  $n_c$ . Then we fix the interaction range by looking at the overall probability of the data as a function of  $n_c$ . In general there is a clear optimum (Fig. 2B), from which we finally infer the maximum entropy value of both parameters,  $n_c$  and  $J$ . We repeat this procedure for every snapshot of each flock and compute the mean and standard deviation of the interaction parameters for each flock over time. Alternatively, for a given flock we can average the log-likelihood over many snapshots, and then optimize, and this procedure gives equivalent results for  $J$  and  $n_c$  (see [SI Appendix, section V](#)).

In Fig. 2C and D we report the value of the interaction strength  $J$  and of the interaction range  $n_c$  for all flocks, as a function of the flock’s spatial size,  $L$ . The inferred values of  $J$  and  $n_c$  are reproducible, although error bars are larger for smaller flocks. In particular,  $J$  and  $n_c$  do not show any significant trend with the flocks’ linear dimensions, with the number of birds, or with the density. This result is not obvious, nor is it in any way built in to our framework; for example, if the real interactions extended over long distances, then our fitting procedure would produce an increase of  $n_c$  and  $J$  with the size of the flock.

In Fig. 2E we also show that the interaction range  $n_c$  does not depend on the typical distance between neighboring birds,  $r_1$ , which is closely related to the flock’s density. Of course, we can



**Fig. 2.** Setting the strength and range of interactions. (A) The predicted strength of correlation,  $C_{\text{int}}$ , as a function of the interaction strength  $J$ , with  $n_c = 11$ , for the snapshot in Fig. 1. Matching the experimental value of  $C_{\text{int}} = 0.99592$  determines  $J = 45.73$ . (Inset) Zoom of the crossing point; error bars are obtained from the model's predictions of fluctuations of  $C_{\text{int}}(J, n_c)$ . (B) The log-likelihood of the data per bird ( $\langle \ln P(\{\vec{s}_i\}) \rangle_{\text{exp}}/N$ ) as a function of  $n_c$  with  $J$  optimized for each value of  $n_c$ ; same snapshot as in A. There is a clear maximum at  $n_c = 11$ . (Inset) the log-likelihood per bird for other snapshots of the same flocking event. (C) The inferred value of  $J$  for all observed flocks, shown as a function of the flock's size. Each point corresponds to an average over all the snapshots of the same flock. Error bars are standard deviations across multiple snapshots. (D) As in C but for the inferred values of  $n_c$ . Averaging over all flocks we find  $n_c = 21.2 \pm 1.7$  (black line). (E) The inferred value of the topological range  $n_c^{-1/3}$  as a function of the mean interbird distance in the flock, for all flocks. Error bars are standard deviations across multiple snapshots of the same flock. (F) As in E but for the metric range  $r_c$ . If interactions extend over some fixed metric distance  $r_0$ , then we expect  $n_c^{-1/3} \propto r_1/r_0$  and  $r_c = \text{constant}$ ; we find the opposite pattern, which is a signature of interactions with a fixed number of topological neighbors (14).

run exactly the same method using a metric interaction range,  $r_c$ , rather than a topological range,  $n_c$ . We simply set  $J_{ij} = J$  if and only if birds  $i$  and  $j$  lie within  $r_c$  meters. In this way we find that the metric range  $r_c$  does depend on the nearest neighbor distance  $r_1$ , in contrast with the topological range  $n_c$  (Fig. 2F). This result provides strong support for the claim put forward in ref. 14 that birds interact with a fixed number of neighbors, rather than with all the birds within a fixed metric distance.

**Model Predictions.** Having fixed  $J$  and  $n_c$  by matching the scalar correlation in the flock, we have no free parameters—everything that we calculate now is a parameter-free prediction. We begin by computing the correlations between pairs of birds as a function of their distance,  $C(r) \sim \sum_{ij} \vec{s}_i \cdot \vec{s}_j \delta(r - r_{ij})$ , as shown in Fig. 3A. There is extremely good agreement across the full range of distances. As we have seen, our maximum entropy calculation finds local interactions, i.e., a relatively small value of  $n_c$  ( $n_c \sim 20$  for flocks of up to thousands birds). This result implies that the scalar correlation  $C_{\text{int}}$ , used as an experimental input to the calculation, is the integral of  $C(r)$  only over a very small interval close to  $r = 0$ : Only the average value of pair correlations at very short distances is used as an input to the calculation, whereas all the long range part of  $C(r)$  is not. Nevertheless, we have very good agreement out to the overall extent of the flock itself. This finding confirms our expectation that a model for the local correlations is able to describe correlations on all length scales. We draw attention to the fact that the apparent correlation length, defined by  $C(r = \xi) = 0$ , is predicted to scale with the linear size of the flock ( $\xi \propto L$ , Fig. 3A, Inset), as observed experimentally (17).

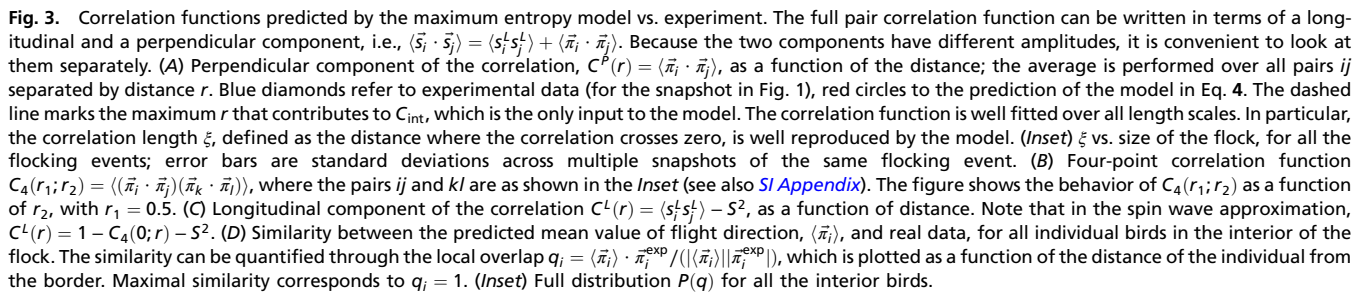
Correlations exist not just between pairs of birds, but among larger  $n$ -tuplets. In Fig. 3B we consider the correlations among quadruplets of birds. Although these correlations are small, their shape is nontrivial and quite noiseless. The model, which takes only local pairwise correlations as input, reproduces very accurately these four-body correlations, including a nonmonotonic

dependence on distance, out to distances comparable to the full extent of the flock. Again, we are not making a fit but a parameter-free prediction.

Finally, instead of measuring correlations, we can ask the model to predict the actual flight directions of individual birds in the interior of the flock, given the directions chosen by birds on the border. This prediction can't work perfectly, because within the model individual birds have an element of randomness in their choice of direction relative to their neighbors, but as shown in Fig. 3D the overlap between predicted and observed directions is very good, not just for birds close to the border but throughout the entire “thickness” of the flock. This result shows that the inference procedure and the model predictions work remarkably well for all individuals in the group. For a discussion of variability across different snapshots and flocks see *SI Appendix*.

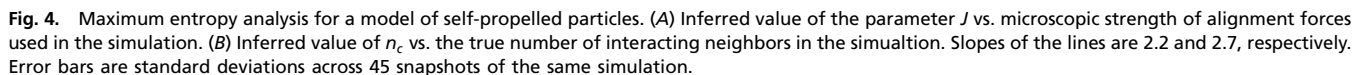
**Testing the Mechanistic Interpretation.** The maximum entropy model has a mechanistic interpretation, from Eq. 3, in terms of social forces driving the alignment of the flight directions. Given the success of the model in predicting the propagation of order throughout the flock, it is interesting to ask whether we can take this mechanistic interpretation seriously. As a test, we have simulated a population of self-propelled particles in three dimensions moving according to social forces that tend to align each particle with the average direction of its neighbors, as described by Eqs. 9 and 10 in *Materials and Methods*. We then compared the simulation parameters ( $J^{\text{sim}}, n_c^{\text{sim}}$ ) to the values ( $J^{\text{mem}}, n_c^{\text{mem}}$ ) obtained by applying the maximum entropy method to snapshots drawn from the simulation, just as we have analyzed the real data. Both the strength and the range of the interaction given by the maximum entropy analysis are proportional to the “microscopic” parameters used in the simulation (Fig. 4A and B), although the maximum entropy interaction range  $n_c^{\text{mem}}$  is roughly  $3 \times$  larger than the true number of interacting neighbors,  $n_c^{\text{sim}}$ . We believe that this overestimation is due to the fact that birds (unlike spins) move through the flock, encountering new neighbors





To summarize, we have constructed the minimal model that is consistent with a single number characterizing the interactions among birds in a flock, the average correlation between the flight directions of immediate neighbors. Perhaps surprisingly, this provides an essentially complete theory for the propagation of directional order throughout the flock, with no free parameters. The theory predicts major qualitative effects, such as the presence of long range, scale-free correlations among pairs of birds, as well as smaller, detailed effects such as the nonmonotonic distance dependence of (four-point) correlations among two pairs of birds. The structure of the model corresponds to pairwise interactions with a fixed number of (topological) neighbors, rather than with all neighbors that fall within a certain (metric) distance; the relevant number of neighbors and the strength of the interaction are remarkably robust across multiple flocking events.

For a long time, theoretical studies of collective animal behavior have relied on arbitrary (albeit reasonable) modeling



assumptions. Recently, thanks to more sophisticated experimental techniques, it has finally been made possible to directly fit the models' parameters to experimental data and perform model selection (19–21). Our work, though, differs from current model fitting approaches in two respects. First, as mentioned in the introduction, we use a general guiding principle (maximum entropy) that spares us from the uneasiness of making assumptions. Second, we consider large groups of individuals, up to several thousands starlings, to be compared with a maximum of few tens (with a typical size of four) analyzed in the lab. Hence, we derive a quantitative theory for the large scale behavior, where the collective phenomenon is fully blown. At the moment, we do not know to what extent the interactions inferred at the level of few individuals are linearly “scalable” up to the thousands, so it seems important to formulate a theory directly at the collective level.

Our approach can be seen as part of a larger effort using maximum entropy methods to link the collective behavior of real biological systems to theories grounded in statistical mechanics (22–34). As with these other examples, we view the success of our theory as an encouraging first step. We have focused on the flight directions, taking the positions of the birds as given. A full theory must connect the velocities of the birds to their evolving positions, which requires more accurate measurements of trajectories over time, and we must consider the fluctuations in the speed as well as direction of flight. There are maximum entropy approaches to both of these problems, and the mapping from maximum entropy models to statistical mechanics suggests that the observation of scale-free correlations in speed fluctuations (17) will locate models of flocking at an especially interesting point in their parameter space (35).

## Materials and Methods

**Data.** Analyzed data were obtained from experiments on large flocks of starlings (*Sturnus vulgaris*), in the field. Using stereometric photography and innovative computer vision techniques (12, 13) the individual 3D coordinates and velocities were measured in cohesive groups of up to 4,268 individuals (11, 14, 17). The dataset comprises 21 distinct flocking events, with sizes ranging from 122 to 4,268 individuals and linear extensions from 9.1 m to 85.7 m (see *SI Appendix, Table S1* for details). Each event consists of up to 40 consecutive 3D configurations (individual positions and velocities), at time intervals of 1/10 s. All events correspond to very polarized flocks (polarization between 0.844 and 0.992; see *SI Appendix, Table S1*). Given a flock at a given time, its exterior border can be computed using appropriate triangulation algorithms (see ref. 13 and *SI Appendix*). Boundary birds are then defined as those who belong to the exterior border.

**Analytic Approach to the Maximum Entropy Model.** To apply the maximum entropy analysis, we need to compute the expected values of correlation functions according to the measure defined in Eq. 1. To this end we must first compute the partition function  $Z(\{J_{ij}\})$ , which is, in general, a hard task. Flocks, however, are very ordered groups, in that birds' velocities are neatly aligned to each other (17). In this case we can use the “spin wave” approximation (36), which exploits the strong ordering condition. Let us call  $\vec{S} = (1/N)\sum_i \vec{s}_i = S\vec{n}$  the global order parameter, or polarization, measuring the degree of collective alignment, where  $\vec{n}$  is a unit vector. Individual orientations can be rewritten in terms of a longitudinal and a perpendicular component:  $\vec{s}_i = s_i^L \vec{n} + \vec{s}_i^\perp$ . If the system is highly polarized,  $S \sim 1$ ,  $|\vec{s}_i^\perp| \ll 1$ , and  $s_i^L \sim 1 - |\vec{s}_i^\perp|^2/2$ ; we verify this last condition for our data in *SI Appendix, Fig. S1*. The partition function can be written as an integral over the  $\{\vec{s}_i\}$ , and if  $S \sim 1$  the leading terms are (see *SI Appendix, section II*, for details):

$$Z(\{J_{ij}\}) \approx \int d^N \vec{\pi} \exp \left[ -\frac{1}{2} \sum_{i,j=1}^N A_{ij} \vec{\pi}_i \cdot \vec{\pi}_j + \frac{1}{2} \sum_{i,j=1}^N J_{ij} \right], \quad [6]$$

where  $A_{ij} = \sum_k J_{ik} \delta_{ij} - J_{ij}$ ,  $d^N \vec{\pi} = \prod_i d\vec{\pi}_i$ , and the  $\{\vec{s}_i\}$  satisfy the constraint  $\sum_i \vec{s}_i = 0$ . If we consider the flight directions of birds on the border as given, integration must be performed with respect to internal variables only (see *SI Appendix, section IV*). After some algebra one gets

$$Z(\{J_{ij}\}; \mathcal{B}) = \int d^{\mathcal{J}} \vec{\pi} \exp \left[ -\frac{1}{2} \sum_{i,j \in \mathcal{J}} A_{ij} \vec{\pi}_i \cdot \vec{\pi}_j + \sum_{i \in \mathcal{J}} \vec{\pi}_i \cdot \vec{h}_i + \frac{1}{2} \sum_{i,j \in \mathcal{J}} J_{ij} + \frac{1}{2} \sum_{i \in \mathcal{J}} h_i^L + \frac{1}{2} \sum_{i,j \in \mathcal{B}} J_{ij} \vec{s}_i \cdot \vec{s}_j \right]. \quad [7]$$

Here  $\mathcal{J}$  and  $\mathcal{B}$  represent the subsets of, respectively, internal and border individuals;  $\vec{h}_i = \sum_{j \in \mathcal{B}} J_{ij} \vec{s}_j$  is a ‘field’ describing the influence of birds on the border on internal bird  $i$ ; and, now,  $A_{ij} = \delta_{ij} (\sum_{k \in \mathcal{J}} J_{ik} + \sum_{l \in \mathcal{B}} J_{il} s_l^L)$ . The integral in Eq. 7 can be carried out explicitly; see *SI Appendix, Eqs. S35–S40*. The reduced model in Eq. 4 corresponds to  $J_{ij} = J n_{ij}$ , with  $n_{ij} = 1, 1/2$ , or 0 according to whether both individuals, just one, or none, belong to the local  $n_c$ -neighborhood of the other. Given the individual coordinates of birds in space, the matrix  $A_{ij}$  can be computed for any given snapshot, and  $Z(J, n_c; \mathcal{B})$  (and correlation functions) can be calculated as a function of  $J$  and  $n_c$ . These two parameters must then be adjusted to maximize the log-likelihood of the data,

$$\langle \log P(\{\vec{s}_i\}) \rangle_{\text{exp}} = -\log Z(J, n_c; \mathcal{B}) + \frac{1}{2} J n_c N C_{\text{int}}^{\text{exp}}. \quad [8]$$

Maximizing with respect to  $J$  corresponds to equating expected and experimental correlations. In our case, this equation can be solved analytically, leading to an explicit expression of the optimal  $J$  vs.  $n_c$ ; see *SI Appendix, Eq. S46*. Maximization with respect to  $n_c$  can then be performed numerically. A graphical visualization of the solution can be found in Fig. 2.

**Self-Propelled Particle Model.** We consider a model of self-propelled particles extensively studied in the literature (9). Each particle moves with vector velocity  $\vec{v}_i(t)$  according to the following equations:

$$\vec{v}_i(t+1) = v_0 \Theta \left[ \alpha \sum_{j \in n_c^i} \vec{v}_j(t) + \beta \sum_{j \in n_c^i} \vec{f}_{ij} + n_c \vec{\eta}_i \right] \quad [9]$$

$$\vec{x}_i(t+1) = \vec{x}_i(t) + \vec{v}_i(t), \quad [10]$$

where  $\Theta$  is a normalization operator  $\Theta(\vec{y}) = \vec{y}/|\vec{y}|$  that serves to keep the speed fixed at  $|\vec{v}| = v_0$ , and  $j \in n_c^i$  means that  $j$  belongs to the  $n_c$  interacting neighbors of  $i$ . The distance-dependent force  $\vec{f}_{ij}$  acts along the direction connecting  $i$  and  $j$ ; following ref. 9, if  $\vec{e}_{ij}$  is the unit vector between  $i$  and  $j$ , we take

$$\vec{f}_{ij}(r_{ij} < r_b) = -\infty \vec{e}_{ij}, \quad [11]$$

$$\vec{f}_{ij}(r_b < r_{ij} < r_a) = \frac{1}{4} \cdot \frac{r_{ij} - r_e}{r_a - r_e} \vec{e}_{ij}, \quad [12]$$

$$\vec{f}_{ij}(r_a < r_{ij} < r_0) = \vec{e}_{ij}. \quad [13]$$

Finally,  $\vec{\eta}_i$  is a random unit vector, independent for each bird and at each moment of time. The parameters  $\alpha$  and  $\beta$  tune the strength of the alignment and of the cohesion force, respectively; in particular, the strength of alignment is given by  $J = v_0 \alpha / n_c$ . To test the maximum entropy analysis, we modified the model in such a way that we could vary  $n_c$ . Specifically, we introduced an angular resolution  $\mu$  such that only neighbors with mutual angles larger than  $\mu$  were included in the neighborhood. When  $\mu$  is of the order of the Voronoi angle the model is statistically equivalent to the original version (where Voronoi neighbors were considered), but increasing (decreasing)  $\mu$  one can decrease (increase) the value of  $n_c$ . In this way both the number  $n_c$  of interacting neighbors and the strength of the interaction  $J$  can be arbitrarily tuned. Parameters were chosen as  $r_0 = 1$  (to set the scale of distance),  $r_b = 0.2$ ,  $r_e = 0.5$ ,  $r_a = 0.8$ ,  $\alpha = 35$ ,  $\beta = 5$ ,  $v_0 = 0.05$ , and we simulated a flock of  $N = 512$  birds. Additional simulations were run (see *SI Appendix, section IX* and Fig. S7) to check that the metric dependency of the distance-dependent force does not affect the relationship between inferred and real values of  $J$  and  $n_c$ .

**ACKNOWLEDGMENTS.** We thank I. Couzin, T. Grigera, N. Leonard, G. Parisi, G. Theraulaz, and J. Toner for helpful discussions. Work in Princeton was supported in part by National Science Foundation Grants IIS-0613435 and PHY-0957573; work in Rome was supported in part by Grants IIT-Seed

Artswarm, ERC-StG n. 257126 and AFOSR-Z809101. Our collaboration was facilitated by the Initiative for the Theoretical Sciences at the Graduate

Center of the City University of New York, supported in part by the Burroughs-Wellcome Fund.

- Camazine S, et al. (2003) *Self-Organization in Biological Systems* (Princeton Univ Press, Princeton).
- Anderson PW (1972) More is different. *Science* 177:393–396.
- Wilson KG (1979) Problems in physics with many scales of length. *Sci Am* 241:158–179.
- Couzin ID, Krause J (2003) Self-organization and collective behavior in vertebrates. *Adv Study Behav* 32:1–75.
- Binney J, Dowrick NJ, Fisher AJ, Newman MEJ (1992) *The Theory of Critical Phenomena* (Oxford Univ Press, Oxford).
- 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* 75:1226–1229.
- Toner J, Tu Y (1995) Long-range order in a two-dimensional dynamical XY model: How birds fly together. *Phys Rev Lett* 75:4326–4329.
- Toner J, Tu Y (1998) Flocks, herds, and schools: A quantitative theory of flocking. *Phys Rev E* 58:4828–4858.
- Grégoire G, Chaté H (2004) Onset of collective and cohesive motion. *Phys Rev Lett* 92:025702.
- Jaynes ET (1957) Information theory and statistical mechanics. *Phys Rev* 106:620–630.
- Ballerini M, et al. (2008) Empirical investigation of starling flocks: A benchmark study in collective animal behaviour. *Anim Behav* 76:201–215.
- Cavagna A, et al. (2008) The STARFLAG handbook on collective animal behaviour: 1. Empirical methods. *Anim Behav* 76:217–236.
- 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.
- 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* 105:1232–1237.
- Cocco S, Monasson R (2011) Adaptive cluster expansion for inferring Boltzmann machines with noisy data. *Phys Rev Lett* 106:090601.
- Cocco S, Monasson R, Sessak V (2011) High-dimensional inference with the generalized Hopfield model: Principal component analysis and corrections. *Phys Rev E* 83:051123.
- Cavagna A, et al. (2010) Scale-free correlations in starling flocks. *Proc Natl Acad Sci USA* 107:11865–11870.
- Cavagna A, et al. (2010) From empirical data to inter-individual interactions: unveiling the rules of collective animal behaviour. *Math Models Methods Appl Sci* 20:1491–1510.
- Lukeman R, Li Y-X, Edelstein-Keshet L (2010) Inferring individual rules from collective behavior. *Proc Natl Acad Sci USA* 107:12576–12580.
- Katz Y, Tunström K, Ioannou CC, Huepe C, Couzin ID (2011) Inferring the structure and dynamics of interactions in schooling fish. *Proc Natl Acad Sci USA* 108:18720–18725.
- Herbert-Read JE, et al. (2011) Inferring the rules of interaction of shoaling fish. *Proc Natl Acad Sci USA* 108:18726–18731.
- Schneidman E, Berry MJ, Il, Segev R, Bialek W (2006) Weak pairwise correlations imply strongly correlated network states in a neural population. *Nature* 440:1007–1012.
- Tkačik G, Schneidman E, Berry MJ, Il, Bialek W (2006) Ising models for networks of real neurons. arXiv:q-bio/0611072.
- Shlens J, et al. (2006) The structure of multi-neuron firing patterns in primate retina. *J Neurosci* 26:8254–8266.
- Lezon TR, Banavar JR, Cieplak M, Maritan A, Federoff NV (2006) Using the principle of entropy maximization to infer genetic interaction networks from gene expression patterns. *Proc Natl Acad Sci USA* 103:19033–19038.
- Tkačik G (2007) Information flow in biological networks. Dissertation (Princeton University, Princeton, NJ).
- Bialek W, Ranganathan R (2007) Rediscovering the power of pairwise interactions. arXiv:0712.4397 [q-bio.QM].
- Tang A, et al. (2008) A maximum entropy model applied to spatial and temporal correlations from cortical networks in vitro. *J Neurosci* 28:505–518.
- Yu S, Huang D, Singer W, Nikolić D (2008) A small world of neural synchrony. *Cereb Cortex* 18:2891–2901.
- Tkačik G, Schneidman E, Berry ML, Il, Bialek W (2009) Spin glass models for networks of real neurons. arXiv:0912.5409 [q-bio.NC].
- 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* 106:67–72.
- Halabi N, Rivoire O, Leibler S, Ranganathan R (2009) Protein sectors: Evolutionary units of three-dimensional structure. *Cell* 138:774–786.
- Mora T, Walczak AM, Bialek W, Callan CG (2010) Maximum entropy models for antibody diversity. *Proc Natl Acad Sci USA* 107:5405–5410.
- Stephens GJ, Bialek W (2010) Statistical mechanics of letters in words. *Phys Rev E* 81:066119.
- Mora T, Bialek W (2011) Are biological systems poised at criticality? *J Stat Phys* 144:268–302.
- Dyson FJ (1956) General theory of spin-wave interactions. *Phys Rev* 102:1217–1230.