

Stochastic dynamics of order and disorder in milling whirligig beetles

Jitesh Jhavar^{1,2,3,4,*}, Vishwesha Guttal^{3,†} and Shashi Thutupalli^{5,6,†}

¹Max Planck Institute of Animal Behaviour, 78465 Konstanz, Germany

²University of Konstanz, 78465 Konstanz, Germany

³Centre for Ecological Sciences, Indian Institute of Science, Bengaluru 560012, India

⁴School of Arts and Sciences, Ahmedabad University, Ahmedabad 380009, India

⁵Simons Centre for the Study of Living Machines, National Centre for Biological Sciences, Tata Institute for Fundamental Research, Bengaluru 560065, India

⁶International Centre for Theoretical Sciences, Tata Institute for Fundamental Research, Bengaluru 560089, India



(Received 8 April 2024; accepted 11 November 2025; published 17 December 2025)

The dynamics of polar order have received much attention in the literature on collective motion. However, milling dynamics, where organisms exhibit rotational motion around the group center, is understudied. Here, we investigate the stochastic dynamics of milling swarms of whirligig beetles [*Dineutus discolor* (Coleoptera: Gyrinidae)] in small to medium groups (50 to 200) where noise arising due to the finite sizes of the swarms cannot be ignored. We quantify the milling order parameter of beetle swarms and show that they exhibit a “transition” from a disordered state to a milling state as a function of group size. We employ a method of data-driven discovery of dynamical equations to obtain stochastic differential equations (SDEs) that disentangle deterministic and stochastic factors driving the milling dynamics. Crucially, we find that the milling order at large group sizes can be explained largely by the deterministic terms of the SDE, while intrinsic stochasticity switches the system between clockwise and counterclockwise milling states. Furthermore, we find that an agent-based model in which each agent aligns with the average direction of its local neighbors and attracts toward neighbors on a larger spatial scale produces stochastic dynamics of milling (and corresponding SDEs) that are consistent with the data-derived SDEs. However, contrary to expectations from mean field theory, both data and explicit agent-based models show that the dynamical equations can change their qualitative features with the group size even when individual interaction rules do not change. In summary, we show evidence for the milling order transition and characterize the role of intrinsic stochastic dynamics in swarming beetles.

DOI: [10.1103/kmbg-sqqc](https://doi.org/10.1103/kmbg-sqqc)

I. INTRODUCTION

Stochasticity is inherent to biological collectives at all scales—from molecular networks [1] and cells [2] to ecosystems [3]. While the properties of collectives have been well studied over the last few decades [4–7], the structure of the noise that governs the collective dynamics has received little attention (however, see Refs. [4,8–11]). Furthermore, among various forms of collective motion, much of the focus has been on group polarization, or the polar order parameter, which measures the degree of directional synchrony among individuals [12,13]. However, milling, a form of collective motion that resembles a circular or vortexlike motion and is observed in many taxa, from insects to fish schools to tadpoles, has

received relatively less attention [14–17]. In this paper, we characterize the stochastic dynamics of the emergent milling motion of whirligig beetle swarms. To do this, we use a data-driven dynamical system framework and a spatially explicit agent-based model that reproduces key features observed in the data.

Typically, theories of collective motion involve stochastic dynamical equations for a group property (a “coarse-grained” variable) [18,19]. These equations are often derived based on assumptions about the microscopic behaviors of organisms or equations written based on some phenomenological rules, such as symmetry considerations [4,20–22]. Using a data-driven equation discovery approach, it is possible to test such theories by constructing the functional forms of the governing dynamical equations directly from the analysis of fluctuations in the state variable, with minimal inputs from the user [23–25]. Such data-driven equations can enable us to disentangle the roles of deterministic and stochastic factors in driving collective motion. Moreover, although the equations are constructed based on the properties of fluctuations at the finest timescale data, the model can accurately capture the long-term dynamics and steady-state distributions of the state variable [25]. Recently, several studies have employed a data-driven approach to construct governing stochastic dynamical

*Contact author: jitesh.jhavar@ahduni.edu.in

†These authors contributed equally to this work.

Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI. Open access publication funded by Max Planck Society.

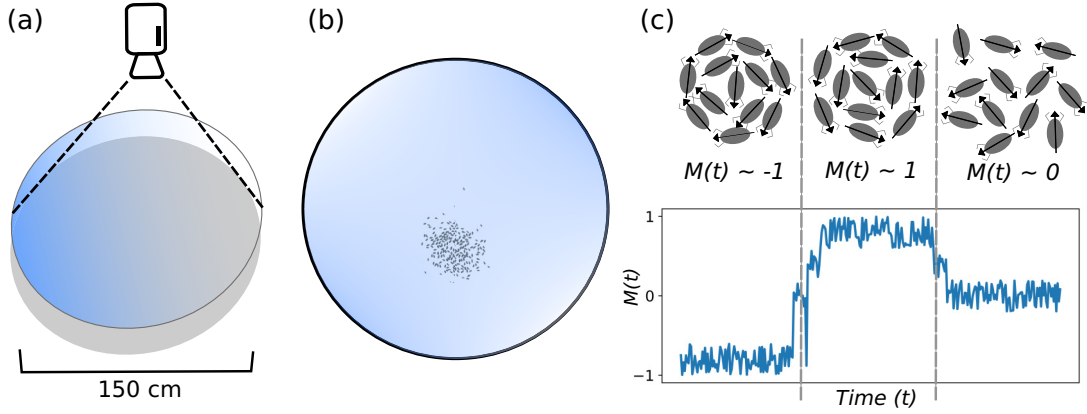


FIG. 1. Experimental setup and the milling order parameter. (a) Schematic representation of the experimental setup consisting of a water-holding experimental tank of 150 cm diameter and an overhead camera for video recording the beetles. (b) A snapshot from one of the experimental trials ($N = 200$), representative of how the beetles self-organize into a tight cluster, typically not close to the boundary. All the group sizes studied generally tend to be 1–2 group lengths away from the tank boundary (see Video S1 in the Supplemental Material [30]). (c) Angular polarization of the group ($M(t)$) is the average angular velocity (normalized) of individuals about the group centroid and ranges between ± 1 . The values of $M(t) \approx -1$ when nearly all individuals move in a clockwise direction about the group centroid and $M(t) \approx 1$ when they all move counterclockwise. For no net angular motion, $M(t) \approx 0$. The stochastic temporal dynamics of the group behavior is calculated from individuals' movement data generated using automated detection and tracking procedures.

equations from data and models [8,9,24–27]. However, these studies have primarily focused on analyzing group polarization as a coarse-grained variable.

In this context, we investigate the collective dynamics of swarms of whirligig beetles [*Dineutus discolor* (Coleoptera: Gyrinidae)], which readily exhibit milling or circular group motion in our experimental setting. Using empirical data from these freely moving groups, we analyze how finite group sizes influence the stochastic nature of their collective behavior. Specifically, we aim to derive the effective equations describing milling, how they depend on the group size, and whether the empirical results can be understood within the framework of existing theories of collective motion.

We summarize our key findings here: First, we find that beetle swarms exhibit swarming at low group sizes and transition to a milling order at higher group sizes. Using the data-driven equation discovery approach [25], we then derive stochastic differential equations (SDEs) directly from empirical time series of milling order. We find that the deterministic part of the dynamics is sufficient to explain the observed order, while stochasticity of finite swarm size leads to transitions between clockwise and counterclockwise milling order. Interestingly, we find that the qualitative form of the deterministic part of the governing equations changes with group size. We then show that the Couzin model [28], one of the classic agent-based models in which individuals align with the motion of nearby neighbors and are attracted to others on a longer spatial scale, reproduces the empirically obtained SDEs, including the group-size-dependent behavior even when the underlying behavioral rules remain the same. In contrast, a mean-field analysis of the model alone would have led to the different conclusion that the interactions among beetles themselves change with group size. Together, these analyses reveal a disorder-to-milling transition in beetle swarms and demonstrate that data-driven equation discovery provides a powerful framework for connecting

empirical observations with mechanistic models of collective motion.

II. EXPERIMENTAL SETUP AND ORDER PARAMETER

Whirligig beetles were placed in a circular arena with a diameter of 150 cm, filled with water to a depth of 30 cm, as previously described in Ref. [29]. We conducted trials for three group sizes ($N = 50, 100$, and 200), with five replicates for each group size with no individual being reused between trials. Using an overhead camera [Fig. 1(a)], we recorded each trial for about 5 min at 30 frames per second (FPS).

We observed that the beetles readily formed swarms, and since the beetles move on the water's surface, the swarms were effectively two dimensional. Throughout the experimental period, individuals within the group continuously moved, but the swarm displayed little translational motion. The swarms were largely away from the arena's walls, thereby minimizing any interactions of the swarm with the boundaries [Fig. 1(b)]. Most importantly, the swarm displayed milling—a swarming state in which many individuals spontaneously move around the center of the group (see Movie S1 in the Supplemental Material [30]).

We measured the positions of individuals for each video frame (t) using our customized code for animal detection and tracking of individuals across frames written in MATLAB [31]. For details, including codes and data, see the Data Availability section, Appendix A, and Movies S2 and S3 in the Supplemental Material [30] for visualization of experiments [30]. We then quantified the milling degree of the swarms using their angular polarization (M) as an order parameter, defined as

$$M(t) = \frac{1}{N} \sum_{i=1}^N \Omega_i(t), \quad (1)$$

where $\Omega_i(t)$ is the instantaneous normalized angular velocity of the i th individual, given by

$$\Omega_i(t) = \frac{\vec{r}_i(t) \times \vec{v}_i(t)}{|\vec{r}_i(t)| |\vec{v}_i(t)|}, \quad (2)$$

where $\vec{r}_i(t)$ is the individual position vector from the centroid of the group of the i th individual. $\vec{v}_i(t)$ is the instantaneous i th individual's velocity. By the above definition, $\Omega_i \in [-1, 1]$ and $\Omega_i = 1$ for counterclockwise motion and $\Omega_i = -1$ in the clockwise direction. $\Omega_i = 0$ for direction toward or away from the group center. The scalar angular polarization $M(t) \in [-1, 1]$ is an average of the individual Ω values. If all the individuals move in the counterclockwise (or clockwise) direction, then $M(t) = 1$ ($M(t) = -1$), implying perfect milling. Finally, $M(t) \approx 0$ implies the lack of milling order in the system [see Fig. 1(c) for an illustration].

Some remarks on the temporal scales of the system are in order. The beetle velocity is computed using the positions $\vec{x}_i(t)$ for $i \in [1, N]$ for each frame based on movement between two consecutive frames ($\delta t = 1/30$ s). The milling order parameter exhibits an autocorrelation time (τ_c) on the order of 10 s, calculated in the following sections. Thus, the time intervals used to compute velocities are two orders of magnitude smaller than the correlation time of the milling order parameter (M). This enables us to capture fine-scale temporal fluctuations in the order parameter arising from stochastic individual-level behaviors.

III. ANGULAR POLARIZATION INCREASES WITH INCREASING GROUP SIZE

In Figs. 2(a)–2(c), we display sample frames of detection and tracking. We observe that the density of swarming beetles increases as a function of group size, from $N = 50$ to $N = 200$. Visual inspection also reveals that the group is more structured or “ordered” for larger groups (Movie S1 in the Supplemental Material [30]).

The above observations are quantified by the milling order parameter ($M(t)$). As shown in Figs. 2(d) and 2(e), $M(t)$ fluctuates around a mean value of ≈ 0 for $N = 50$ and 100, exhibiting a unimodal distribution around $M \approx 0$ for all trials of $N = 50$ and 3 out of 5 trials for $N = 100$ [Figs. 2(g) and 2(h)]. Thus, the beetle swarms exhibit a disordered state for smaller groups with no net group angular polarization.

On the other hand, for $N = 200$, the order parameter predominantly exhibits negative values, representing a clockwise milling group motion of the beetle swarms. However, fluctuations in this state continue to occasionally result in a transition to the counterclockwise milling state [Fig. 2(f)]. This results in an asymmetric bimodal distribution of the order parameter [Fig. 2(i)], with a higher frequency observed for the clockwise milling state. More specifically, four out of the five trials show two modes at $|M| \approx \pm 0.3$, with a larger peak at $M \approx -0.3$ and a smaller one at $M \approx 0.3$. One trial, however, shows a single mode at $M \approx -0.3$ [green color line in Fig. 2(i)]. This asymmetry could arise due to a shorter experimental duration or an inherent group tendency to rotate in one direction more than the other. In the following sections, we further analyze this aspect using the spatial model.

IV. MILLING IS DETERMINISTIC ORDER WITH NOISE-DRIVEN TRANSITIONS

The observed increase in order with group size and the bimodal distribution for $N = 200$ are an intriguing feature of the beetle swarms. Such a bimodality in the order parameter can be seen in dynamical systems characterized by three fixed points, where two ordered states correspond to stable fixed points that are separated by an unstable intermediate fixed point (which may correspond to a disordered state), and an additive noise, i.e., a constant variance stochasticity, drives spontaneous transitions between the two ordered stable states [25,32]. Such a system can be thought of as a deterministic ordered system with noise-induced switching between different ordered states. Alternatively, a counterintuitive mechanism involving multiplicative noise can lead to bimodality even if the underlying deterministic dynamical system is characterized by a single stable disordered state. In other words, the observed ordered state can be a noise-induced state that is absent in the deterministic counterpart of the model. This phenomenon has been shown in both theory [11,19] and experiments of collective animal motion [27].

This leads us to the following question: Is the observed bimodality in milling a consequence of deterministic order with stochastic switching causing bimodality, or is it due to noise-induced bimodality? To address this, we take the approach of data-driven stochastic equation discovery, which enables the construction of governing stochastic dynamical equations, described in detail below.

To mathematically describe and thereupon investigate the underlying structure of the beetle swarm dynamics, following the jump-moments methods described in Appendix B, we derive the following stochastic differential equation from the empirical time-series data of M :

$$\dot{M} = F(M) + G(M) \eta(t), \quad (3)$$

where η represents uncorrelated Gaussian white noise, i.e., $\langle \eta(t) \rangle = 0$ and $\langle \eta(t) \eta(t') \rangle = 0 \forall t, t' \in [0, T]$ [33]. This Langevin dynamical framework, interpreted in the Itô sense [33], assumes that the underlying noise structure is Gaussian and uncorrelated. We test these assumptions using our diagnostics of the data-derived model in Appendix C.

Before proceeding further, we emphasize that Langevin equations for order parameter dynamics are typically constructed via a coarse-graining analytical procedure starting from a microscopic model of interactions (e.g., [11,19]). However, we take an inverse modeling or a data-driven equation discovery approach of deriving the Langevin equation directly from the high-resolution time-series data [11,25]. Here, we emphasize that this method is not merely a fitting procedure to find the coefficients of a known functional form. Rather, the data-driven discovery approach allows us to infer the form of the drift and diffusion functions directly from stochastic time-series analysis. We then later find the best polynomial fit using the sparse regression method [25]. In the process, we reveal how the functional forms of drift and diffusion change with beetle group size. Later, in the next section, we employ a suite of collective behavior models—spatial and nonspatial—to capture the patterns of the data-derived

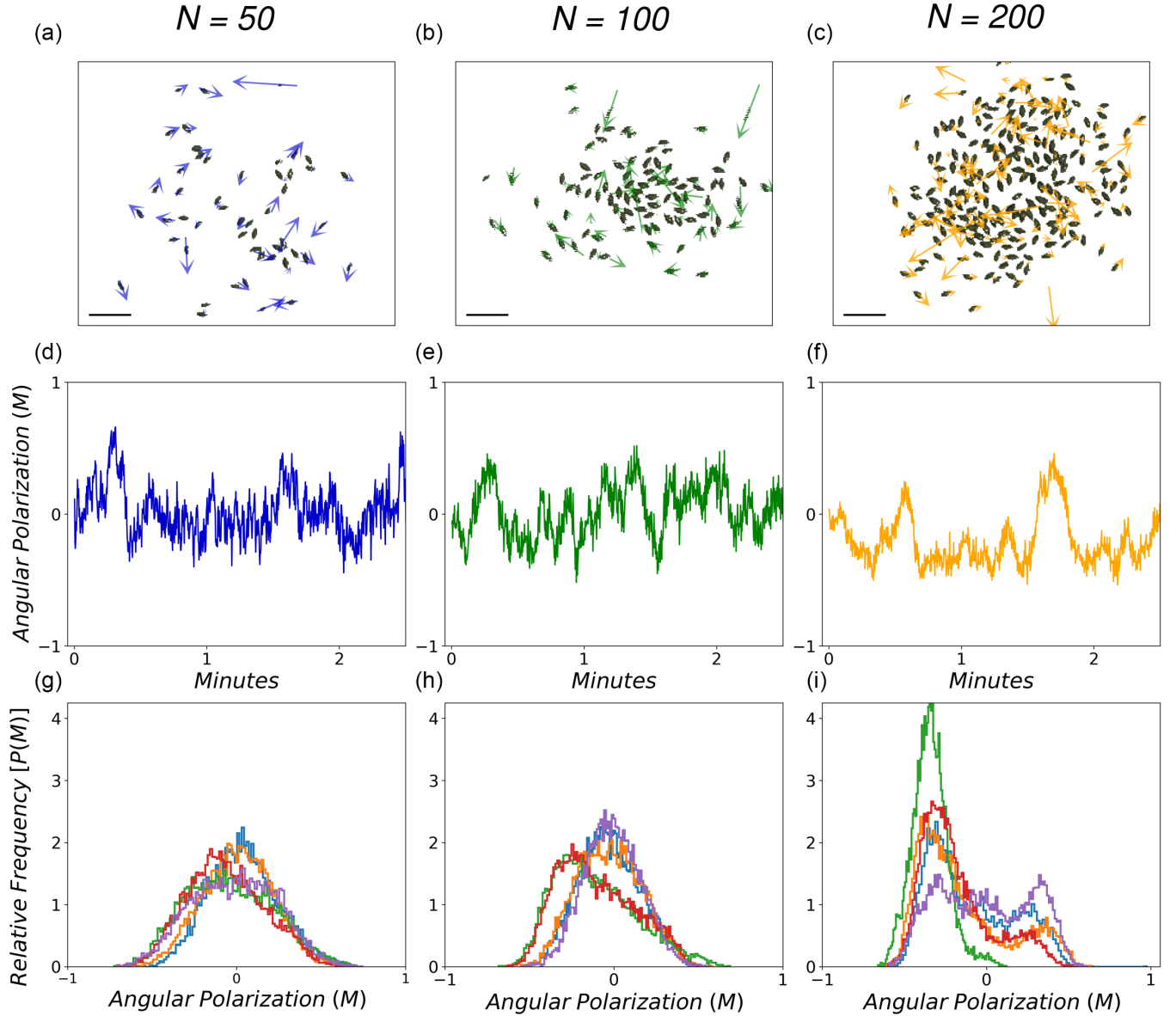


FIG. 2. Angular polarization increases with group size. (a)–(c) Sample snapshots of individuals depicting their group behavior for different group sizes. Arrows are overlaid using data from our automated tracking procedures, and the length of the arrow indicates the individual's velocity. The scale bars at the bottom left in all the panels represent 5 cm of length in the tank (which has a diameter of 150 cm). (d)–(f) Sample time series of $M(t)$. For $N = 50$, M remains closer to 0, whereas for $N = 100$, M shows greater variation but still varies around 0. For $N = 200$, M varies around finite values, representing milling order, and occasionally transitions between positive and negative values. (g)–(i) Probability distributions of $M(t)$ from all experimental trials. We see a unimodal distribution at $M \approx 0$ for $N = 50$ and 100 and a bimodal distribution at $M \approx \pm 0.3$ for $N = 200$. The shifting of the modes from 0 to finite M with group size demonstrates that angular polarization or milling increases with group size in whirlingig beetles.

SDE and individual-level rules that best account for the data-derived Langevin dynamics of the collective motion.

Data-derived SDE

We derive the deterministic function by sparse regression using the PyDaddy package [25] (Appendix B). Comparison with the conventional method of regression to the binned data of coefficients [34–36] did not lead to any qualitative differences. For the stochastic term, we fit an inverted parabolic function using methods from Refs. [34–36], considering the analytical expectations of individual-based models

(Appendix D). We show the results in Fig. 3 and describe them as follows.

$N = 50$. The deterministic and the stochastic parts $F(M)$ and $G(M)$ of the Langevin dynamics in Eq. (3) are shown in Figs. 3(a) and 3(d). The functional form of the equation obtained by fitting a cubic drift and a quadratic diffusion function to the data is

$$\dot{M} = [-2.155M + 1.743M^2 - 2.534M^3] + \sqrt{[0.23 - 0.17M^2]} \eta(t), \quad (4)$$

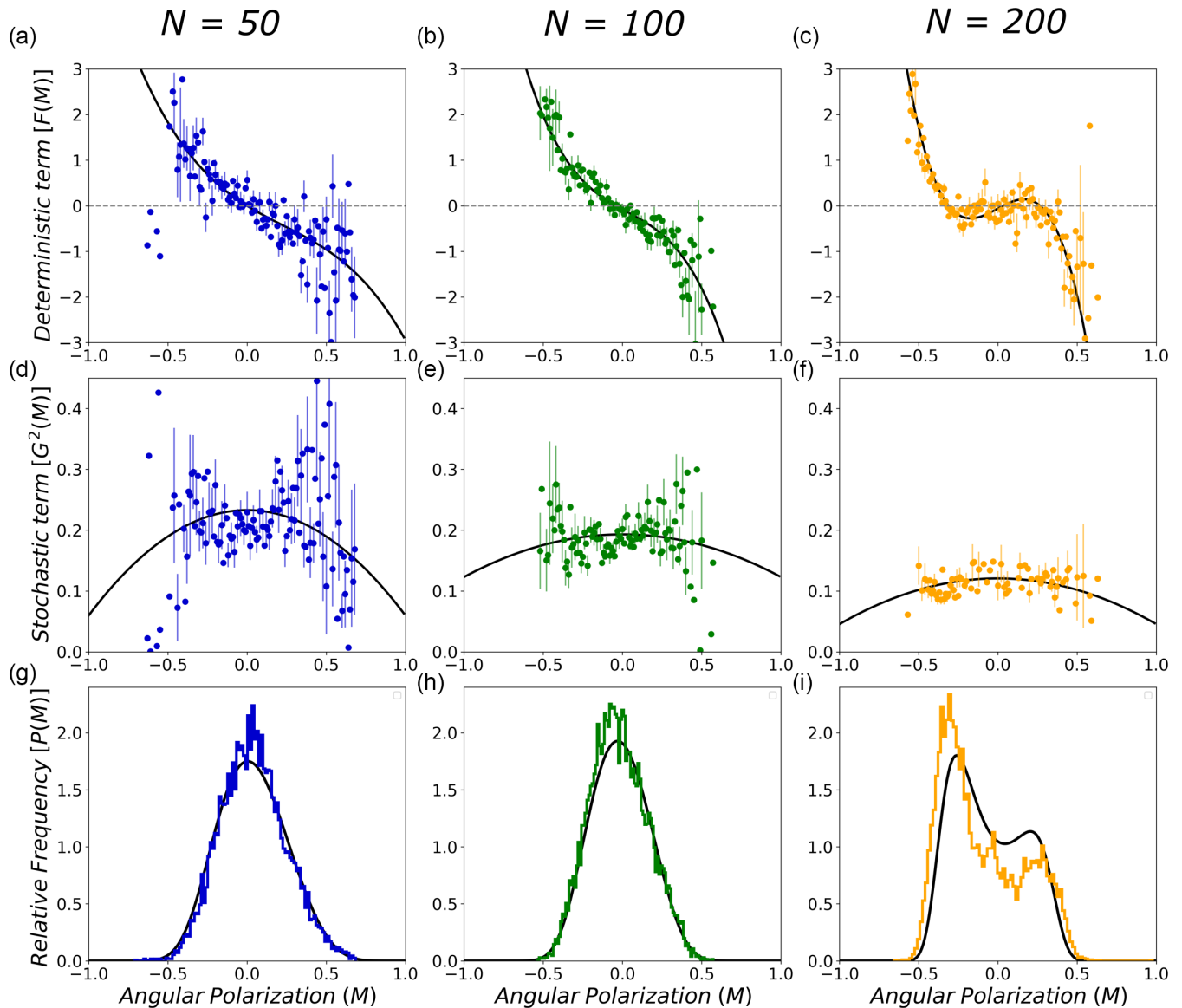


FIG. 3. Data-derived stochastic differential equations of the form $\dot{M} = f(M) + G(M)\eta(t)$ [Eq. (3)] from the experimental time-series data. (a)–(c) For all the group sizes, the deterministic component ($F(M)$) of the data-derived SDE varies as a cubic function of angular polarization (M). The strength of the cubic term increases with group size [panels (a) to (c)]. Furthermore, the number of fixed points also changes with the group size and switches their stability [panels (a) to (c)]. Group sizes $N = 50$ [panel (a)] and 100 [panel (b)] both show one stable fixed point at $M^* = 0$. The group size $N = 200$ [panel (c)] exhibits two stable fixed points at finite $M^* \approx \pm 0.3$, while the disorder ($M^* = 0$) is now an unstable fixed point. (d)–(f) The stochastic component ($G^2(M)$) of the SDE shows a weak multiplicative form as a function M with overall strength of stochasticity decreasing with group size. Black curves in the first two rows are the individual fits to the functions obtained by minimizing the least-squares distance (L^2) between the data and the fitted functions, which together form the data-derived SDE. The error bars in the plots of deterministic and stochastic coefficients represent the standard error of F and G^2 for a given value of M . (g)–(i) The relative frequency plots are obtained by merging all the trials for different group sizes studied, whereas the black curves are the analytical solution of the data-derived SDE, demonstrating self-consistency of the data-derived model.

where $\eta(t)$ is Gaussian white noise. $F(M)$ exhibits a stable fixed point at $M^* \approx 0$ (i.e., $\frac{dF}{dM}|_{M^* \approx 0} < 0$). It is characterized by a weak cubic function ($R^2 = 0.60$). Graphical inspection reveals that the stochastic part $G(M)$ is weakly multiplicative; i.e., the strength of stochasticity is slightly greater for values of M close to 0 than at the extremities of the data. We note that the lack of data beyond $M \sim \pm 0.5$ leads to a possibility of considering an additive fit to $G^2(M)$; Fig. 3(d) shows a plot of $G^2(M)$ with the fitted

curve. We consider a quadratic fit to $G^2(M)$ —a parsimonious functional form expected by analytical derivations of simple collective behavior models that account for intrinsic stochasticity [11] (Appendix D). In Fig. 3(g), we show that the analytically derived probability density function (PDF), from the data-derived SDE, matches the PDF of M for $N = 50$. We refer to Appendix E for a derivation and the formula for the general form of the SDE we obtained from data. We determine the $P_s(M)$ using this derivation, find the

normalization constant by finding the definite integral under the appropriate limits, and then numerically plot the final expression of $P_s(M)$.

$N = 100$. The data-derived equation is given by

$$\dot{M} = [-1.887M + 0.635M^2 - 7.574M^3 - 0.078] + \sqrt{[0.19 - 0.07M^2]\eta(t)}. \quad (5)$$

We observe that the overall patterns for $N = 100$ for the deterministic and the stochastic parts are qualitatively similar to those obtained for $N = 50$. However, for the deterministic part F , the cubic term ($R^2 = 0.81$) has a larger coefficient [Fig. 3(b)], and the overall strength of the stochastic part G is lower [Fig. 3(e)]. Finally, the analytically derived steady-state PDF (see Appendix E) also matches with the data [Fig. 3(h)].

$N = 200$. The data-derived SDE for this group size is given by

$$\dot{M} = [1.868M - 0.251M^2 - 22.21M^3 - 0.059] + \sqrt{[0.12 - 0.07M^2]\eta(t)}. \quad (6)$$

Here, the cubic nature of the deterministic part ($R^2 = 0.80$) F is further evident. The function exhibits multiple crossings of the x axis [Fig. 3(c)], suggesting alternating stable and unstable fixed points. More parsimoniously, the deterministic dynamics has three fixed points: $M^* \approx \pm 0.3$ and $M^* = 0$. $M^* \approx \pm 0.3$ are stable and $M^* = 0$ is an unstable fixed point. The stochastic part G is similar to its counterparts from the smaller group sizes ($N = 50$ and 100) and is weakly multiplicative, with an overall strength that is least when compared to the other group sizes [Fig. 3(f)].

Due to multiple stable fixed points in the deterministic part of the SDE of $N = 200$, the dynamics is qualitatively different from the other two group sizes. The distribution of M from all merged trials retains the bimodal nature from most trials. Furthermore, the analytically derived steady-state PDF (see Appendix E) from our data-derived equation is, visually, quite close to the observed distribution [Fig. 3(i)].

Furthermore, we also confront our data-derived SDEs by comparing the autocorrelation functions (ACFs) from their simulated time series with those from the data (Appendix C). We find that both the shape of the ACFs from the data and the model and the timescales of their decay are roughly similar (~ 10 s). Hence, the SDE model passes the self-consistency test for its applicability to the data dynamics, thereby reassuring the accuracy and predictive power of our data-derived SDE for the milling beetle swarms across all group sizes.

Put together, from Eqs. (4)–(6) and Figs. 2 and 3, the dynamics of angular polarization M can be summarized as follows:

(1) The deterministic part F can be well approximated as a cubic polynomial showing an increase in its cubic strength with increasing group size. For smaller group sizes, there is only one stable fixed point at $M^* = 0$ (corresponding to disorder) in the dynamics. For the largest group size, however, $M^* = 0$ becomes unstable, and we find two stable fixed points at $M \approx \pm 0.3$ corresponding to finite order. Hence, we refer to the milling order in the system as deterministic.

(2) The stochastic part G is weakly multiplicative. For lower values of M , i.e., disorder, the strength of stochasticity

TABLE I. Spatial model parameters.

Parameter	Symbol	Value
Speed	s	5
Radius of repulsion zone	R_r	1
Radius of orientation zone	R_o	9.0
Radius of attraction zone	R_a	22
Turning angle constraint	T_c	0.06π
Blind angle	Bl	0.85π
Noise	η	1.0
Length of arena	L	100
Time interval	dt	0.1
Group size	N	50, 100, 200

is large, while for higher values of M (i.e., ordered states), the strength of stochasticity is lower. The overall strength of the multiplicative noise decreases with increasing group size N . This decreasing strength of the stochastic term or the intrinsic noise with group size N is in coherence with theoretical expectations as the variance is, generally, expected to decay as $1/\sqrt{N}$ (Appendix D).

(3) We infer that the combination of deterministic order and stochasticity yields spontaneous switches between stable clockwise and counterclockwise order in milling beetles; however, although we infer that the noise is multiplicative, the multiplicative nature is too weak to cause any major alteration in this dynamics qualitatively, such as producing higher states of order [27].

V. A SPATIALLY EXPLICIT COLLECTIVE MOTION MODEL REPRODUCES KEY FEATURES OF THE DATA

Our aim now is to arrive at a potential mechanism of interactions between individuals that best explains the observed milling dynamics of beetle swarms: (1) increasing milling order as a function of group size and (2) SDEs for the milling order parameter that are qualitatively consistent with the data-derived SDEs. To do so, we consider a spatially explicit model of collective behavior from literature by Couzin *et al.* [28]. In this model, agents form a group and exhibit a milling state when the individuals interact via a local-range alignment (i.e., align the direction of motion with the average direction of neighbors over a short distance) and a longer-range attraction (i.e., move toward the average/centroid of positions of neighbors present over a longer distance).

We simulate this model for group sizes, $N = 50, 100$, and 200 , to compare the three group sizes considered in our experiments (see Appendix F for details of the model). While varying N , we keep the rest of the parameters constant to generate the model time series for a regime exhibiting milling (Table I). As the model has many parameters, milling can occur for a wide set of parameter combinations, even within the range of N that we consider. Therefore, our goal is solely to demonstrate that the model can capture the experimental dynamics, at least within the chosen parameter space. Using the same procedure for milling data, we analyze the time series, frequency distributions, and

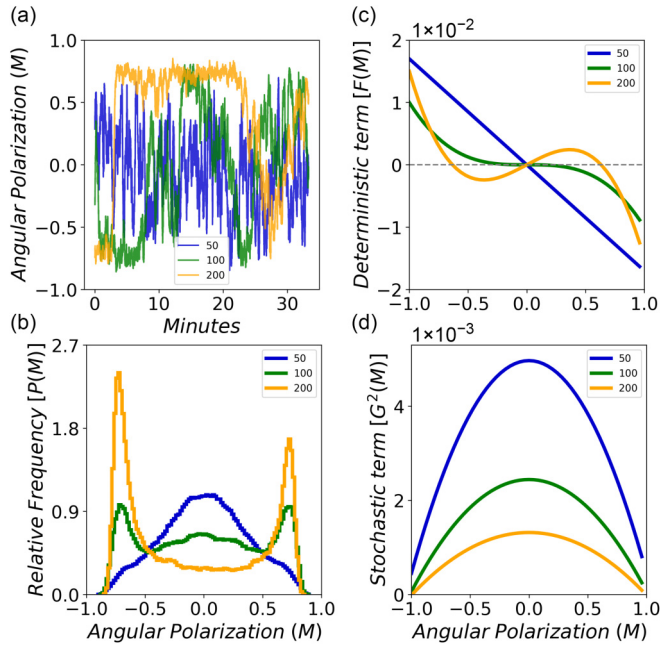


FIG. 4. Spatially explicit agent-based model captures key qualitative features of experimental dynamics, including the SDE of the milling order parameter. (a) The instantaneous milling order parameter M is plotted as a function of time for different group sizes from the model. (b) The distribution of M for various group sizes. With increasing N , the switch from unimodal to bimodal behavior is consistent between the data and the spatial model. (c) The deterministic functional form of $F(M)$ obtained using the package PyDaddy [25] that uses sparse regression technique. The strength of the cubic term of F increases with group size from $N = 50$ to 200, from 0 to a finite value. Additional stable fixed points arise for larger N , while the stability of the $M = 0$ state changes from stable to unstable. (d) The stochastic term $G^2(M)$ is a multiplicative function of M . The strength of G^2 is more for the disordered state, i.e., low M , and less for the ordered state, i.e., high M . Moreover, the overall strength of G^2 decreases with N .

the SDE of the simulated data of milling order parameter M from the spatially explicit model (Fig. 4).

Figure 4(a) shows the time series of M from the model for the three N . While for smaller group sizes it broadly fluctuates between $M \approx -0.5$ and 0.5 , for $N = 200$, M remains in the high-ordered states ($M \approx \pm 0.7$). For the intermediate group size, the system shows multistability between the swarming and the milling states. Note that individual trials do show multiple states for intermediate group size ($N = 100$) in our data [Fig. 2(h)].

The distribution plots of M for various group sizes clearly show an increase in the angular order with group size as depicted by Fig. 4(b), which is consistent with our experimental data. Notably, although there is no behavior built in the model that can produce uneven modes in the distribution, the asymmetry in the distributions (for larger N) is retained even for the time series of a million time steps run of the model. This result from the model suggests that the asymmetry in our experimental data [Fig. 3(i)] is potentially a stochastic effect due to short time series instead

of a systematic bias of the beetles to rotate more in one direction.

A. Stochastic differential equation of the spatial model

The transition from disorder to order with group size in the spatial model matches our data and is suggestive that the group-level dynamics of the spatial model is also deterministically ordered dynamics. To ascertain this, we derive the stochastic differential equation of the dynamics from the agent-based model. We derive the SDE of the model in a similar manner as for the experimental data (Sec. IV). Below, we write the data-derived equation from the model data to compare them with its counterpart from the experimental data [Eqs. (4)–(6)],

$$\dot{M} = \alpha M - \beta M^3 + \left[\frac{\gamma(1 - M^2)}{N} \right]^{1/2} \eta(t), \quad (7)$$

where $\alpha \approx -0.02$, 0 , and 0.01 , $\beta \approx 0$, 0.01 , and 0.025 , for, respectively, and $\gamma \approx 0.2 \forall N$. Figures 4(c) and 4(d) show the functional fits to the data of F and G^2 for $N = 50, 100$, and 200 . Strikingly, like our data, Fig. 4(c) shows that the deterministic term $F(M)$ changes as a function of group size. While the smaller group size ($N = 50$) exhibits a stable fixed point at $M = 0$, the larger group size ($N = 200$) exhibits three fixed points of which $M = 0$ becomes unstable, but the fixed points at finite M are stable. Therefore, the broad qualitative feature of group-size-dependent deterministic function, from linear to cubic, of the data is captured by our spatially explicit model [Figs. 3(a)–3(c)].

Furthermore, the stochastic term [Fig. 4(d)] $G^2(M)$ is multiplicative, consistent with the data [Figs. 3(d)–3(f)]. In addition, the strength of G^2 decreases with increasing group size, as also in the data. Notably, in both the data and the model, the overall strength of G^2 decreases approximately by a factor of 2 with doubling the group size [Figs. 3(d) and 3(f)] or is proportional to $1/N$ [Eq. (7)]. Strikingly, this feature from the spatial model also corroborates with the analytical form of the SDE from the individual-based nonspatial models (Appendix D), where the strength of G^2 also turns out to be proportional to $1/N$.

Together, we find that both the dynamics and the stochastic differential equation from the spatial model qualitatively match the salient features that we analyze in our experimental data on the collective motion of whirling beetles. In addition, we also quantify the mean residence time distribution between the clockwise and counterclockwise milling states, both in the data and in the model (Appendix G). The analysis mainly describes that the distribution of the mean residence time in a given state of order is exponential, and the strength of the decay decreases with N ; i.e., the system stays in highly ordered states more often in the larger groups. Hence, through this analysis, we quantitatively describe the nature of transitions and show that the spatial model also captures the nature of these stochastic transitions in our data. Finally, we note that although the noise ($G^2(M)$) has a multiplicative form, it does not qualitatively alter the deterministic states, as in the case of polarized schools of fish [27]. However, the noise is essential for stochastic transitions between the different ordered states of M .

Nonspatial individual-based models

In addition to the spatial model in the previous subsection, we also analyze a mean-field or nonspatial modeling framework where analytical derivation of the SDE, using a coarse-graining procedure, is possible [11,19,21]. In this framework, we assume that an individual can exhibit two discrete angular directions, i.e., clockwise or counterclockwise. Individuals can randomly switch between the two directions or copy others' directions. At any instant, agents are clockwise (denoted by -1) or counterclockwise (denoted by $+1$). The average angular direction of all individuals gives the order parameter (m) in this model. Hence, this order parameter has the same features as the order parameter we constructed from real data and the spatially explicit model.

Under these assumptions, we can analytically derive the stochastic differential equation of m as a function of individual-level reaction rates. We describe the details of this derivation in the Supplemental Material [30] and Appendix D) and also point to Refs. [11,19,21] for similar derivations in the context of group polarization.

Subsequently, we compare the analytical equations from a suite of individual-based model (IBMs) from this framework to the coefficients of the SDE of data. To achieve this, we use the least-squares method to fit the functional form of model SDE to the data and assess the goodness of fit using the adjusted R^2 , which provides the fitted individual-level behaviors in the form of reaction rates. Importantly, the analytically derived model SDEs show that the deterministic part of the SDEs does not depend on the group size N (Appendix D). Hence, to fit the SDEs from data, where the deterministic function changes qualitatively with group size, the fitted model's individual-level rate parameters vary with group size. In other words, we find that the strength of the higher-order ternary interaction increases with the group size, which is required to capture the increase in the cubic degree of the deterministic function with N in the data. This interpretation differs from the spatially explicit model, in which we can explain the group-size-dependent deterministic function without altering the individual-level parameters across different group sizes. Therefore, we argue that the spatially explicit model offers a more parsimonious explanation for the observed data.

VI. DISCUSSION

In summary, our data-derived SDEs reveal that the observed milling order at larger group sizes corresponds to a deterministic stable fixed point; additionally, stochastic fluctuations arising from small swarm size induce transitions between the stable clockwise and stable counterclockwise milling states. Furthermore, we demonstrate that a spatially explicit agent-based model, where individuals follow a direction determined by the average directional alignment of their neighbors, as well as a tendency to be attracted toward the average position of their neighbors over a relatively larger spatial scale than the scale of the directional alignment, reproduces the qualitative features of the data. Our data-driven equation discovery approach is different due to the fact that dynamical equations are derived from data with minimal interference from the researcher. Importantly, the key input to the equation

discovery process is stochastic changes in the order parameter at the finest time scale ($\approx 1/10$ th of a second) available from our observations. Despite that, the data-derived equation can reproduce the long-time steady-state distribution of the order parameter consistent with the data.

Although milling is found in many species, from swarms of insects to schools of fish, this phenomenon is relatively less studied (but see Refs. [16,17,37–40]). In our system, first, we find that the degree of milling of a beetle swarm increases with group size (Fig. 2). Note that we also analyze boundary effects as a potential confounding factor in our results (Appendix H) and find them to be absent or minimally affecting the observed patterns. Next, we find that the functional form of the drift (i.e., the deterministic component of the data-derived SDE) changes qualitatively with group size—from a linear function for small group sizes to a cubic function for medium to larger group sizes (Fig. 3). The deterministic function has a stable fixed point near zero order for smaller group sizes. On the other hand, for a larger group size ($N = 200$), the stable fixed point is located at a finite order. Hence, the order in the system can be explained through the changing form of deterministic dynamics with group size. Recalling the analytical form of the mean-field equations for individual-based collective behavior models [11], we note that this is an interesting departure from the theory; however, recent work by Martin *et al.* [41] provides a theoretical underpinning for this aspect. In the mean-field approximations of individual-based models, the functional form of the deterministic component is independent of the group size [11,19,21]. Would this also be true for spatially explicit models? To explore this, we pursued a spatially explicit agent-based model of collective behavior from the literature [28].

The model of collective motion by Couzin *et al.* [28] assumes that organisms exhibit local averaging of neighbors' movement behavior. More specifically, the model allows for milling motion when agents follow the average alignment of their neighbors over a small scale, while also exhibiting a tendency to stay cohesive by interacting with neighbors over a larger scale [28,37,42]. Through our comparison, we find that the degree of milling increases as a function of group size in both the data and the spatially explicit model (Fig. 4). Moreover, the model not only captures the functional forms of both the deterministic and the stochastic dynamics of milling but crucially also captures the changes in the deterministic functional form with group size (Fig. 4); i.e., the deterministic function changes from a linear function at small group sizes to a nonlinear function (which may be approximated as a cubic function) at larger group sizes, qualitatively similar to as observed in the data-derived SDE of milling dynamics. We emphasize that the spatial model captures the group-size dependence of the deterministic function even though the interactions between agents were fixed and did not change with the group size. In contrast, if we had relied only on mean-field models, our conclusion would have been different, i.e., interactions among beetles change with group size (Appendix D). Thus, averaging interactions among neighboring individuals in a spatial model could be a more parsimonious explanation that can produce the studied group-level dynamics of the whirling beetle swarms. At a more microscopic level, the explanation from the spatial model is somewhat

similar to the explanation of the milling motion of daphnia swarms under the influence of some external environment [15]. The similarities between the two explanations, *vis-à-vis*, are an attraction force, an avoidance force, and an asymmetric perception. However, in the *Daphnia* model, the attraction is toward a global attractor, whereas in the spatial model, the attraction is somewhat more local.

Looking beyond milling, in other collective systems, similar interactions, such as local averaging or strong feedback between individuals, have been used to describe behaviors in collectives, including bees [43], ants [44], and locusts [9,45]. However, such higher-order or averaging interactions are not always necessary to produce collective behavior; for example, simpler interactions, such as copying the direction of one randomly chosen neighbor, could explain the collective motion in other systems like fish [27,46,47], fruit flies [48], and midges [49].

For any emergent pattern of a group, many different local-level behaviors can produce qualitatively similar properties of larger-scale collective motion. Therefore, inverse methods that decipher individual-level rules from observing group dynamics may have degenerate solutions. This limitation also applies to our method. Our study best illustrates this by contrasting inferences one can draw from the mean-field individual-based model and the spatially explicit agent-based model. We note that both of these minimal models were able to broadly capture the observed dynamics. However, there were interesting contrasts. While the mean-field model required us to invoke the assumption that individual-level interaction strengths or rates change with group size, this was not necessary for the spatially explicit model. Therefore, in such situations, the parsimonious explanation is the preferred one that the spatially explicit model provides; it captures the changing terms of the SDE with group size, even when the interactions are not altered by group size.

Others have studied individual-level interactions using methods like building force maps and measuring pair correlations between individuals [46,50–53]. These analyses typically rely on measuring correlations between individuals that often limit the system size, as it becomes difficult to identify causal relationships when many individuals get involved. In contrast, our method relies on the analysis of fine-scale fluctuations in coarse-grained variables such as group linear or angular order [8]. In addition, our method also provides a means to analyze stochasticity in the system, which can often lead to interesting, counterintuitive transitions or states [9,19,27,54]. Broadly, our method enables the analysis of small to intermediate group sizes, providing a complementary approach to characterize and analyze group dynamics and the underlying mechanisms that explain the observed patterns.

To conclude, we characterized the stochastic dynamics of milling beetles and made inferences about the underlying mechanisms by complementing the analysis with an agent-based model. In the future, such analysis using a stochastic differential equation can be extended to further refinements, such as deriving a coupled partial differential equation [23] or even a coupled stochastic partial differential equation [55] from the real data to assess dynamical relationships such as between density and order [56–58]. This can help us better connect with field-theoretical frameworks that model the

collective motion of “active” systems [20,22]. We also encourage studies on functional aspects of collective behaviors, e.g., an analysis of the costs and benefits of an emergent property of collective motion such as milling. A few studies on this aspect of whirlingigs suggest that beetles aggregate to gain antipredation benefits [29,59,60]. However, the competition’s role and the effects of other ecological processes in shaping such behaviors are still under-investigated. We hope that our study inspires further investigations into using this cutting-edge framework, based on the analysis of stochastic dynamics, to gain a deeper understanding of collective systems.

ACKNOWLEDGMENTS

We thank Colin Twomey for the experimental data with the whirlingig beetles and William Romey for help with the collection of the beetles. We thank Daniel Calovi for his suggestions on model implementation. J.J. acknowledges financial support from the Heidelberger Akademie der Wissenschaften under the WIN (Wissenschaftlichen Nachwuchs) program and the Alexander von Humboldt-Stiftung for a postdoctoral fellowship. V.G. acknowledges the Science and Engineering Research Board (SERB), Government of India (Grant No. CRG/2021/003513) and Department of Science and Technology, Government of India, Fund for Improvement of S&T Infrastructure (DST-FIST). S.T. acknowledges support from the Department of Atomic Energy, Government of India, under Project No. RTI4006, and the Simons Foundation (Grant No. 287975).

J.J., V.G., and S.T.; experiments: S.T.; tracking: J.J.; data and model analysis: J.J., V.G., and S.T.; writing: J.J., V.G., and S.T.

The authors declare no competing interest.

DATA AVAILABILITY

All data supporting the findings of this study—including individual positions, velocities, angular velocities, polarization time series, and the analysis codes and movies—have been deposited in a publicly accessible repository [61].

APPENDIX A: INDIVIDUAL DETECTION AND TRACKING

The tracking of individuals was performed to get their instantaneous positions and velocities from the experimental videos. The procedure consists of two steps. In the first step, the positions of all individuals are determined for all the frames in the videos. Upon basic filtering to remove minimal background noise, we used the image subtraction method to perform this task by subtracting the two images with an approximate time interval of 30 frames between them. In other words, from the current frame, we subtract a far-away frame in time so that objects have moved sufficiently, leaving positive and negative values corresponding to the object positions in the current and the far frame. Selecting one of these, i.e., either all positive or all negative, accordingly, and repeating this process for every frame of the data gives us the positions of individuals in all time frames. Using this procedure, individuals were detected as intensity blobs, of which the peak was used to identify the centroid of each blob. To carry

out this work, several functions were used from the Image Processing Toolbox in MATLAB [31]. The next step involved linking each blob from one frame to the next to assign the identities. We used the Kalman filter algorithm [62] to perform the tracking. The velocity of each individual was calculated by subtracting the X and Y positions of the individuals across consecutive frames.

APPENDIX B: CONSTRUCTION OF THE DETERMINISTIC AND THE STOCHASTIC TERMS OF THE SDE

We determine the functional forms of F and G of Eq. (3) using the following formulas [36]:

$$F(M) = \left\langle \frac{M(t + \Delta t) - M(t)}{\Delta t} \right\rangle_{M(t) \in [m, m+\epsilon]}, \quad (\text{B1})$$

$$G^2(M) = \left\langle \frac{(M(t + \delta t) - M(t))^2}{\delta t} \right\rangle_{M(t) \in [m, m+\epsilon]}. \quad (\text{B2})$$

We set $\epsilon = 0.01$ for all our analyses. $\delta t = \Delta t = 1/30$ is the time between consecutive frames in our experimental videos. For the fitting of F , we used the Python package PyDaddy [25], which bypasses the binning and averaging step in the above formulas and uses a sparse regression method for finding the functional forms. However, for G^2 , we relied on the regular fitting procedure to the coefficients obtained using the above formula [Eq. (B2)], as the sparse regression method did not work for G^2 .

APPENDIX C: DATA DIAGNOSTICS

We analyze some statistical features of our time-series data to determine the timescales of the system and verify the assumptions underlying the SDE model derived from the data [25,36]. A standard technique for analyzing time-series data is to measure its autocorrelation as a function of the time difference, or lag, between two data points. For the angular polarization $M(t)$, its autocorrelation is calculated using

$$\text{ACF}(\tau) = \frac{\langle (M(t + \tau) - \langle M(t) \rangle)(M(t) - \langle M(t) \rangle) \rangle}{\langle (M(t) - \langle M(t) \rangle)^2 \rangle}, \quad (\text{C1})$$

where τ is the time lag over which the ACF is calculated. We calculate the ACF for the angular polarization (M) for different group sizes by merging all trials, as shown in Figs. 5(a)–5(c). We find that the ACF monotonically decreases with the τ . It hits zero at a certain time lag. The point where the ACF hits zero for the first time can be described to be the timescale of the system over which correlations exist (ca. 10 s). Since the SDE we derive from the data uses the smallest time difference or consecutive values of M , which is well within the correlation time, using an SDE framework allows us to analyze the fluctuations at small timescales resulting from individual-level stochastic behaviors.

According to the SDE [Eq. (3)], the underlying noise $\eta(t)$ must be (1) uncorrelated and (2) Gaussian. We analyze the structure of the underlying noise in the angular polarization (M) time series of beetle data to test this assumption. Following Refs. [63–65], we measure the autocorrelation of the residual fluctuations [25,36]; this time series is derived from the original time series of M by subtracting the local mean for

a short time window as

$$\eta(t) = M(t) - \frac{1}{T} \int_{t-\frac{T}{2}}^{t+\frac{T}{2}} M(\tau) d\tau, \quad (\text{C2})$$

where T is the window duration. Figures 5(d)–5(f) show this autocorrelation for noise time series from all group sizes. The ACF of the residual noise time series decreases rapidly. Hence, the assumption of the SDE model that the underlying noise is uncorrelated is satisfied.

We test the assumption that the noise is Gaussian. We plot the cumulative density distribution (CDF) of all ΔM for a representative $M_0 = 0$, as shown in Figs. 5(g)–5(i). Since the underlying noise in the dynamics drives the variation in these differences at the fine timescale, their distribution is a good approximation of the underlying noise structure. These distributions qualitatively resemble the CDF of the fitted Gaussian distribution using the mean and standard deviation of the data; data; Kolmogorov-Smirnov (KS) test statistic [66]: $N = 50$: KS statistic = 0.6312, p -value = 0.2720; $N = 100$: KS statistic = 0.9743, p -value = 0.0013; $N = 200$: KS statistic = 0.7905, p -value = 0.0878. Thus, we find that the statistical features of the data's underlying noise support the modeling assumptions of an SDE framework.

Finally, we also confront the data-derived SDE with the data by simulating the SDE numerically using the framework of Itô calculus [33] for a number of time steps similar to the size of the data time series ($\sim 50\,000$ time steps). Thereupon, we find the autocorrelation function of this time series [Figs. 5(j)–5(l)] and compare it with that of the data [Figs. 5(a)–5(c)]. The shape of the ACFs matches between the data and the numerical simulations of the data-derived SDE, and the timescales of the ACFs are similar (~ 10 s).

APPENDIX D: NONSPATIAL INDIVIDUAL-BASED MODELS OF INCREASING COMPLEXITY

In addition to the spatial model in the main manuscript text (Sec. V), we also analyze a mean-field or nonspatial modeling framework where analytical derivation of the SDE, using a coarse-graining procedure, is possible [11,19,21]. In this modeling framework, we interpret milling as the collective behavior of beetles in which each beetle i can be in one of the two discrete states: $\omega_i = 1$ and $\omega_i = -1$, representing counterclockwise and clockwise movement around the group center, respectively. Individuals stochastically switch between these two states due to interactions with other individuals; they may also switch them entirely randomly. The average of individual states ω_i in the group, at a given time, represents consensus or the milling order (denoted $m(t)$) of the group. If all individuals move in the counterclockwise (or clockwise) direction, the milling order parameter m will be high, i.e., close to 1 (or -1). On the other hand, if individuals are evenly split between two angular directions of motion, the system will show no milling order; i.e., m will be close to zero. Due to stochastic interactions and individual state changes, we expect the milling order parameter in the model to fluctuate stochastically over time. Hence, this order parameter has the same features as the order parameter we constructed from real data and the spatially explicit model.

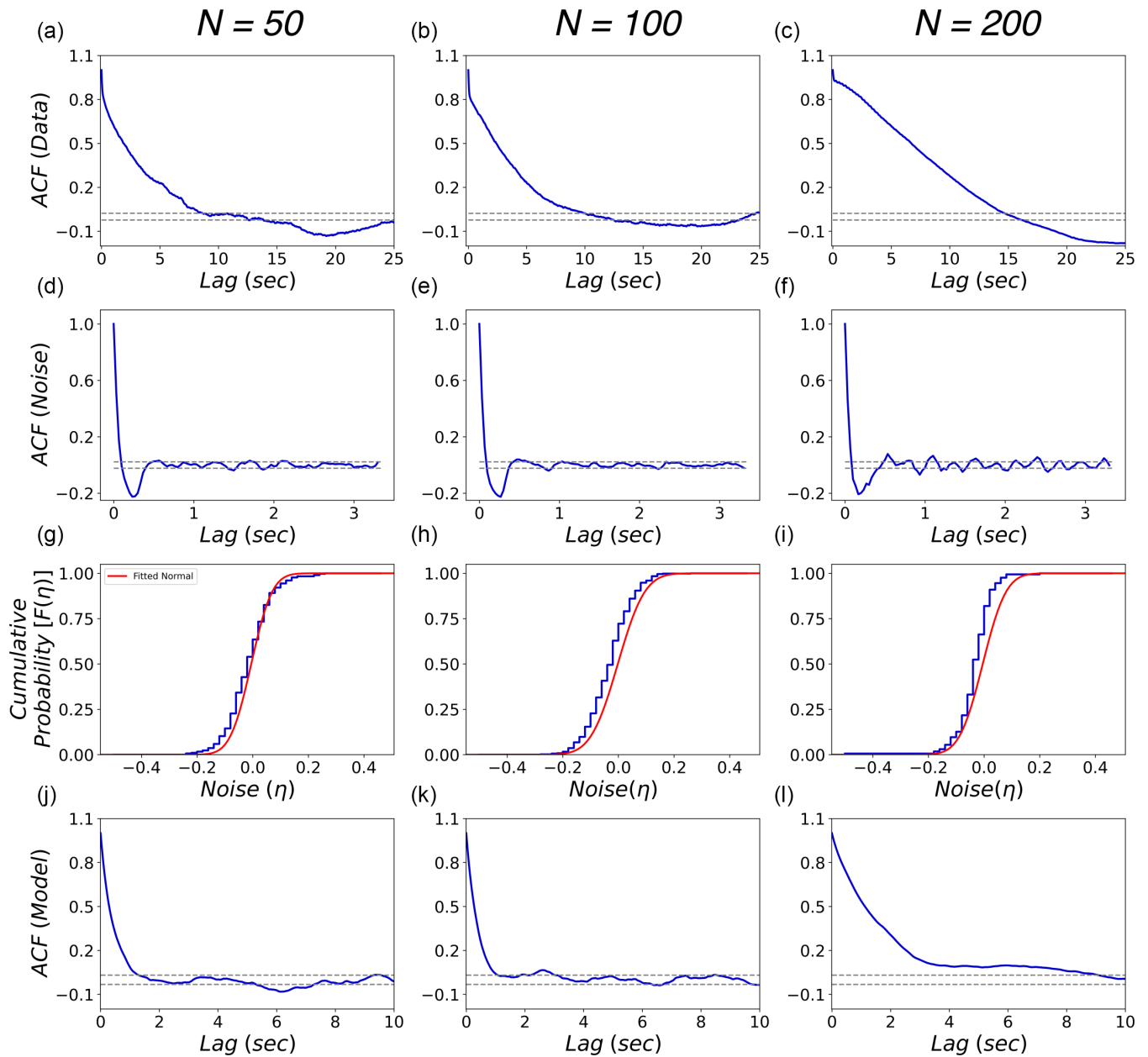


FIG. 5. Test for Gaussian structure in the noise underlying the data. (a)–(c) ACF of the time series of angular polarization M for all group sizes. (d)–(f) The ACF of the time series of noise time series, which is derived from the original time series. The ACF of noise very quickly decays to zero compared to the original data. (g)–(i) The cumulative density distribution of fluctuations from a representative value of M in the data. The red curves represent CDFs of fitted normal function using the norm function from the Python library `scipy.stats`. (j)–(l) The autocorrelation function of the time series from the numerical simulations of the data-derived SDE (3) for the three group sizes in the data.

1. Model rules and SDEs

We follow the conventions of chemical reaction kinetics, where each individual-level interaction occurs stochastically at a certain rate, thus changing the state of the system. In our model, the simplest way an individual can change the angular direction is via random switching, i.e., without any interactions with others. Specifically, we allow for two different spontaneous/random turning rates: (1) from counterclockwise to clockwise, denoted by r_{1-} , and (2) from clockwise to counterclockwise, denoted by r_{1+} . This allows us to detect any potential bias of turning or handedness at the individual level,

which has been observed in some species due to differences in the left and right brain's decision-making rates [67–69].

Further, an individual may interact with only one randomly chosen group member and copy the angular direction of that group member; this is termed stochastic pairwise interaction, which occurs at a rate r_2 (and has been reported in fish [27]). We also consider the possibility that agents interact with more than one group member when they update their direction of motion [46,50]. To achieve this, we allow interactions between up to five randomly chosen group members, each with associated rates denoted by r_3 (for interactions among three members, also known as ternary interactions), r_4 (for

TABLE II. Reactions in the individual-based model.

S. No.	Reaction	Propensity ($a_j(x)$)	ν_{j1}	ν_{j2}
1	$X_1 \xrightarrow{r_1+} X_2$	$r_1 + x_1$	-1	1
2	$X_2 \xrightarrow{r_1-} X_1$	$r_1 - x_2$	1	-1
3	$X_1 + X_2 \xrightarrow{r_2} 2X_1$	$r_2 x_1 x_2$	1	-1
4	$X_1 + X_2 \xrightarrow{r_2} 2X_2$	$r_2 x_1 x_2$	-1	1
5	$2X_1 + X_2 \xrightarrow{r_3} 3X_1$	$r_3 x_1^2 x_2$	1	-1
6	$2X_2 + X_1 \xrightarrow{r_3} 3X_2$	$r_3 x_1 x_2^2$	-1	1
7	$2X_1 \xrightarrow{r_{2s}} X_1 + X_2$	$r_{2s} x_1^2$	-1	1
8	$2X_2 \xrightarrow{r_{2s}} X_1 + X_2$	$r_{2s} x_2^2$	1	-1
9	$3X_1 + X_2 \xrightarrow{r_4} 4X_1$	$r_4 x_1^3 x_2$	1	-1
10	$3X_2 + X_1 \xrightarrow{r_4} 4X_2$	$r_4 x_1 x_2^3$	-1	1
11	$3X_1 + 2X_2 \xrightarrow{r_5} 4X_1 + X_2$	$r_5 x_1^3 x_2^2$	1	-1
12	$3X_2 + 2X_1 \xrightarrow{r_5} 4X_2 + X_1$	$r_5 x_1^2 x_2^3$	-1	1

interactions among four), and so on. While stochastic pairwise interactions are symmetric, i.e., either of the two interacting individuals can change direction, ternary or higher-order interactions can be asymmetric. When the interactions are asymmetric, the individual whose direction differs from the majority switches its direction. For symmetric interactions, the randomly chosen focal individual chooses the direction of motion of the other individual. We note that, from the previous theoretical work, it is known that the polynomial of the drift or the deterministic function has a degree that (roughly) scales with the number of interacting neighbors [11,19,45]. Since the empirically observed degree of the deterministic function polynomial [Fig. 3(c)] appears cubic by visual inspection, we did not consider interactions beyond more than five individuals.

So far, the interactions described can be termed “positive,” i.e., resulting in a directional alignment or copying with the majority. However, there is evidence that organisms in groups may also show “negative” interactions, i.e., repulsive interactions resulting in occasional “splitting” [28,70,71]. Therefore, we also allow for the possibility of negative interactions in our model, but, for simplicity, only in the pairwise interaction rates, denoted by r_{2s} . While other higher-order splitting interactions are, in principle, possible, we do not introduce them in our model for the sake of simplicity.

Based on these interactions, we construct a hierarchy of models with increasing complexity. The simplest model has only two rules—random switching and stochastic pairwise copying, whereas the most complex one has all of the above interactions (see Table II for the list of all models we considered).

For each of these models, using the chemical Langevin method [72], we analytically derive a SDE for the milling order parameter m that links the behavioral interactions with the group-level milling dynamics for analytical derivation. Such a stochastic differential equation is of the form

$$\dot{m} = f(m) + g(m)\eta(t), \quad (D1)$$

where $\eta(t)$ is a Gaussian white noise. In the above equation, $f(m)$, termed as drift, captures the deterministic or average dynamic of the order parameter m . On the other hand, $g(m)$,

referred to as diffusion, captures the stochastic variations. Both these functions depend on the reaction rates described in the model above. We then seek the best fit between the data-derived drift-diffusion functions with the model-derived drift-diffusion functions, considering the above hierarchy of models involving various stochastic interactions. We tested four different models of increasing complexity. The models are denoted by I, II, III, and IV. Model IV consists of all the behavioral rules modeled as chemical reactions (Table II), whereas the other models contain the subsets of these reactions (Table III).

2. Derivation of the macroscopic SDE from the individual-level rules

Our goal is to derive the equation of the following form in terms of the behavioral rules described in Table II:

$$\frac{dm}{dt} = f(m) + g(m)\eta(t). \quad (D2)$$

While a more formal method to derive the macroscopic SDE from the individual-level rules uses the van Kampen system size expansion approach, described in Ref. [21], we use a rather heuristic approach to derive the macroscopic SDE. This approach is known as the chemical Langevin method, introduced by Gillespie [72]. Here, to derive the macroscopic SDE in terms of the individual-level rules, we simply substitute the expressions for the reaction propensity $a_j(x)$ and the state change or the stoichiometric vector ν as described in Table II into the chemical Langevin equation, given by

$$\frac{dx_i(t)}{dt} = \sum_{j=1}^r \nu_{ji} a_j(\mathbf{x}) + \frac{1}{\sqrt{N}} \sum_{j=1}^r \nu_{ji} [a_j(\mathbf{x})]^{1/2} \eta_j(t). \quad (D3)$$

Here, r is the total number of reactions, which is 8 for model III, and N is the total number of individuals in the model. η_j are Gaussian white noise sources with zero mean and correlation $\langle \eta_i(t) \eta_j(t') \rangle = \delta_{ij} \delta(t - t')$. x_i is the total fraction of the individuals in state i . Therefore, the summation of all x_i fractions at any given time is 1. Importantly, the order m is given by $x_1 - x_2$ or $x_2 - x_1$ for a system with two states. Therefore, this simplified framework allows us to relate the dynamics of M derived from data with the dynamics of m from the model.

The deterministic part, which is the rate of change of the population of a given species, is given by

$$\begin{aligned} \sum_{j=1}^r \nu_{j1} a_j(\mathbf{x}) &= \nu_{11} a_1(\mathbf{x}) + \nu_{21} a_2(\mathbf{x}) + \nu_{31} a_3(\mathbf{x}) \\ &+ \nu_{41} a_4(\mathbf{x}) + \nu_{51} a_5(\mathbf{x}) + \nu_{61} a_6(\mathbf{x}) \\ &+ \nu_{71} a_7(\mathbf{x}) + \nu_{81} a_8(\mathbf{x}). \end{aligned} \quad (D4)$$

Upon substituting from Table II,

$$\begin{aligned} \sum_{j=1}^r \nu_{j1} a_j(\mathbf{x}) &= -r_1 + x_1 + r_1 - x_2 + r_2 x_1 x_2 - r_2 x_1 x_2 \\ &+ r_3 x_1^2 x_2 - r_3 x_1 x_2^2 - r_{2s} x_1^2 + r_{2s} x_2^2. \end{aligned} \quad (D5)$$

TABLE III. Group-size dependent behavioural rates inferred using the individual-based models of increasing complexity.

Group size	Trials	Model	Estimated rates							Adj. R^2	Avg. Adj. R^2	$\overline{D_{KL}}$	$\sigma^2\{D_{KL}\}$
			r_{1+}	r_{1-}	r_2	r_3	r_4	r_5	r_{2s}				
50	A	I	2.89	2.34	6.93	—	—	—	—	0.54	0.6	0.06	0.01
	B		2.93	2.4	7.32	—	—	—	—	0.66		0.11	0.01
	C		2.25	2.22	11.65	—	—	—	—	0.68		0.04	0
	D		3.05	3.73	9.94	—	—	—	—	0.46		0.03	0
	E		2.04	2.27	8.48	—	—	—	—	0.66		0.03	0
	A	II	3.23	2.8	4.06	3.92	—	—	—	0.44	0.42	0.07	0.01
	B		1.32	1.11	3.95	0	—	—	—	0.37		0.09	0.01
	C		1.15	1.15	6.74	0	—	—	—	0.43		0.04	0.01
	D		1.06	1.19	3.03	0	—	—	—	0.21		0.02	0
	E		1.23	1.39	5.63	0	—	—	—	0.66		0.03	0.01
	A	III	0.68	0.24	6.62	3.92	—	—	2.56	0.44	0.52	0.07	0.01
	B		1.48	0.95	8.63	0	—	—	1.48	0.66		0.11	0.01
	C		1.91	1.78	0	15.44	—	—	3.8	0.66		0.04	0.01
	D		1.54	2.25	0	15.75	—	—	4.34	0.47		0.05	0.01
	E		3.34	3.56	0.13	9.67	—	—	0.66	0.38		0.03	0
	A	IV	0.98	0.99	1.31	1.22	1.22	0	0.9	0.59	0.32	0.03	0.004
	B		1	0.97	1.71	1.32	1.30	0	0.94	0.40		0.008	0.001
	C		1.24	1.25	9.3	2.6	2.12	0	1.71	0.70		0.03	0.003
	D		0.95	1.03	2.87	1.34	1.31	0	0.95	0.44		0.02	0.004
	E		1.28	1.56	0	6.83	6.18	16.3	2.47	0.66		0.04	0.005
100	A	I	3.08	3.19	14.91	—	—	—	—	0.67	0.63	0.01	0
	B		2.93	2.98	14.11	—	—	—	—	0.71		0.02	0
	C		1.41	1.39	16.72	—	—	—	—	0.47		0.15	0.01
	D		2.07	2.34	15.1	—	—	—	—	0.65		0.08	0
	E		2.92	3.1	11.73	—	—	—	—	0.67		0.02	0
	A	II	1.06	1.09	4.43	0	—	—	—	0.34	0.44	0.02	0
	B		1.08	1.11	5.05	0	—	—	—	0.37		0.02	0
	C		4.71	4.67	0	18.35	—	—	—	0.57		0.19	0.02
	D		4.87	5.09	0	17.86	—	—	—	0.58		0.09	0.01
	E		1.05	1.09	4.29	0	—	—	—	0.34		0.01	0
	A	III	1.83	1.92	0	21.9	—	—	5.7	0.71	0.68	0.02	0
	B		1.78	1.83	0	20.7	—	—	5.35	0.76		0.02	0
	C		0.11	0.06	1.02	23.1	—	—	5.49	0.57		0.2	0
	D		1.6	1.93	0	21.12	—	—	4.55	0.71		0.03	0
	E		1.63	1.71	0	18.2	—	—	4.69	0.64		0.04	0
	A	IV	1.21	1.39	14.4	2.06	1.42	0	2.43	0.68	0.59	0.018	0.002
	B		1.25	1.37	12.7	2.66	1.89	0	2.45	0.72		0.014	0.001
	C		1.11	1.07	10.09	2	1.71	0	1.26	0.48		0.18	0.015
	D		1.04	1.36	13.07	2.41	1.83	0	1.72	0.66		0.06	0.004
	E		0.98	0.99	2.18	1.27	1.25	0	0.95	0.45		0.01	0.003
200	A	I	1.55	1.8	35.87	—	—	—	—	0.32	0.55	0.46	0
	B		0.66	0.72	18.88	—	—	—	—	0.52		0.35	0.03
	C		1.4	2.54	10.5	—	—	—	—	0.63		0.15	0.01
	D		1.21	1.49	17.82	—	—	—	—	0.66		0.3	0.02
	E		0.69	0.72	20.08	—	—	—	—	0.63		0.14	0.01
	A	II	9.6	9.3	02	36.9	—	—	—	0.42	0.68	0.26	0.01
	B		4.67	4.72	0	19.94	—	—	—	0.78		0.19	0.02
	C		3.5	4.49	0	12.1	—	—	—	0.66		0.05	0
	D		4.95	5.26	0	19.52	—	—	—	0.77		0.1	0.01
	E		5.0	5.12	0	19.0	—	—	—	0.8		0.05	0.01
	A	III	2.81	2.52	0	46.1	—	—	8.61	0.44	0.64	0.26	0.01
	B		1.61	1.67	0	23.9	—	—	3.86	0.55		0.77	0.03
	C		0.88	1.86	0.34	15.22	—	—	3.14	0.67		0.04	0
	D		1.52	1.85	0	23.6	—	—	4.35	0.71		0.08	0.01

TABLE III. (Continued.)

Group size	Trials	Model	Estimated rates							Adj. R^2	Avg. Adj. R^2	$\overline{D_{KL}}$	$\sigma^2\{D_{KL}\}$
			r_{1+}	r_{1-}	r_2	r_3	r_4	r_5	r_{2s}				
	E		1.67	1.73	0	25.2	—	—	4.31	0.83		0.04	0
	A		0.96	1.24	24.1	1.08	0.86	0	1.43	0.24		0.3	0.01
	B	IV	1.35	1.37	4.08	0	24.5	28.8	2.62	0.75	0.65	0.34	0.03
	C		1.01	1.23	0.33	0	19.4	44.1	1.30	0.78		0.83	0.05
	D		0.92	1.19	13.1	2.6	2.16	0	1.16	0.64		0.23	0.02
	E		1.22	1.25	6.74	0	21.8	31.3	2.12	0.85		0.3	0.02

Similarly, the stochastic term is given by

$$\begin{aligned}
& \sum_{j=1}^r v_{j1} [a_j(\mathbf{x})]^{1/2} \eta_j(t) \\
&= v_{11} \sqrt{a_1(\mathbf{x})} \eta_1(t) + v_{21} \sqrt{a_2(\mathbf{x})} \eta_2(t) \\
&+ v_{31} \sqrt{a_3(\mathbf{x})} \eta_3(t) + v_{41} \sqrt{a_4(\mathbf{x})} \eta_4(t) \\
&+ v_{51} \sqrt{a_5(\mathbf{x})} \eta_5(t) + v_{61} \sqrt{a_6(\mathbf{x})} \eta_6(t) \\
&+ v_{71} \sqrt{a_7(\mathbf{x})} \eta_7(t) + v_{81} \sqrt{a_8(\mathbf{x})} \eta_8(t). \quad (\text{D6})
\end{aligned}$$

Upon substituting from Table II,

$$\begin{aligned}
& \sum_{j=1}^r v_{j1} [a_j(\mathbf{x})]^{1/2} \eta_j(t) \\
&= -\sqrt{r_{1+} x_1} \eta_1(t) + \sqrt{r_{1-} x_2} \eta_2(t) \\
&+ \sqrt{r_2 x_1 x_2} \eta_3(t) - \sqrt{r_2 x_1 x_2} \eta_4(t) \\
&+ \sqrt{r_3 x_1^2 x_2} \eta_5(t) - \sqrt{r_3 x_1 x_2^2} \eta_6(t) \\
&- \sqrt{r_{2s} x_1^2} \eta_7(t) + \sqrt{r_{2s} x_2^2} \eta_8(t), \quad (\text{D7})
\end{aligned}$$

where η_1 to η_8 are independent Gaussian noise sources. Since $m = x_1 - x_2$ and $x_1 + x_2 = 1$, we have

$$x_1 = \frac{1+m}{2} \quad \text{and} \quad x_2 = \frac{1-m}{2}. \quad (\text{D8})$$

Substituting Eqs. (D5) and (D7) into Eq. (D3) and using Eq. (D8), we get

$$\begin{aligned}
\frac{dm}{dt} &= \left[\frac{-r_{1+}}{2} (1+m) + \frac{r_{1-}}{2} (1-m) - r_{2s} m \right] \\
&+ \left[\frac{r_3}{4} m(1-m^2) \right] + \frac{1}{\sqrt{N}} \left[r_{1+} + r_{1-} + r_2(1-m^2) \right. \\
&\left. + \frac{r_3}{2} (1-m^2) + r_{2s}(1+m^2) \right]^{1/2} \eta(t). \quad (\text{D9})
\end{aligned}$$

Note. The resultant variance of the sum of two independent random variables is the sum of the original variance of their distributions, i.e., $\sqrt{a} \eta_1 + \sqrt{b} \eta_2 = \sqrt{a+b} \eta$ [33]. The above equation is the macroscopic SDE for the individual-based model III that consists of eight reactions (Tables II and III). For model I, we can set $r_3 = r_{2s} = 0$, and for model II, $r_{2s} = 0$ in the above equation. Finally, for model IV, we can substitute all the terms from Table II into the chemical

Langevin equation (D3) to obtain its macroscopic SDE; however, we avoid writing the expressions for all the models here.

3. Fitting macroscopic SDE from the IBMs to the data

As we derived a macroscopic SDE from simple individual-level rules using the chemical Langevin method for a system of N individuals, a similar equation derived from real data (Fig. 3) can be assumed to represent the rules of interactions among real animals. Hence, a direct comparison between the two equations, from the model and the data, can be used to determine the dominant modes of microscopic interactions in our model system. For this, we optimize the behavioral rate parameters of the model that result in the minimum least-square distance (LSD) between the data-derived functions (Fig. 3) and the model SDE (D9). Note that we consider both the deterministic and the stochastic part of the SDE together for this procedure and minimize the summed LSD (L^2). We fit the four models (I, II, III, and IV) of increasing complexity to the data. In Table III, we put together the values of best-fitted reaction rates and the goodness of fit using adjusted R^2 for both qualitative and quantitative assessments of the match between the data and the models. As found from Table III, models II–IV are comparatively similar, especially for the larger group sizes, as they contribute higher-order interactions. Crucially, the rate of higher-order interactions dominates for larger group sizes.

APPENDIX E: STEADY-STATE SOLUTION OF THE FOKKER-PLANCK EQUATION

We consider the following Langevin equation, interpreted in the *Ito* sense:

$$\dot{M} = A(M) + \sqrt{D(M)} \eta(t), \quad (\text{E1})$$

where $\eta(t)$ is Gaussian white noise. To obtain the steady-state solution for the probability density function $P_s(M)$ for this Langevin equation, we consider the corresponding Fokker-Planck equation [33]

$$\frac{\partial P}{\partial t} = -\frac{\partial}{\partial M} [A(M)P(M, t)] + \frac{1}{2} \frac{\partial^2}{\partial M^2} [D(M)P(M, t)]. \quad (\text{E2})$$

At steady state and under zero probability current conditions, we obtain

$$A(M)P_s(M) - \frac{1}{2} \frac{d}{dM} [D(M)P_s(M)] = 0. \quad (\text{E3})$$

Rearranging terms gives us

$$\frac{d}{dM}(\ln[D(M)P_s(M)]) = \frac{2A(M)}{D(M)}. \quad (\text{E4})$$

Integrating, we obtain the steady-state solution

$$\ln P_s(M) = \int \frac{2A(M)}{D(M)} dM - \ln D(M) + C, \quad (\text{E5})$$

or equivalently, we have [33,73]

$$P_s(M) = \frac{\mathcal{N}}{D(M)} \exp\left(-\int 2\frac{A(M)}{D(M)} dM\right),$$

where $\mathcal{N}$ is the normalization constant.

1. Evaluating the integral

Based on the empirical findings, we consider the following forms of the drift and diffusion function:

$$A(M) = a + bM + cM^2 + dM^3$$

and

$$D(M) = \alpha + \beta M^2.$$

Therefore, we obtain

$$\frac{2A(M)}{D(M)} = \frac{2d}{\beta}M + \frac{2c}{\beta} + \frac{R(M)}{\alpha + \beta M^2},$$

where the remainder function $R(M)$ is

$$R(M) = 2\left(b - \frac{d\alpha}{\beta}\right)M + 2\left(a - \frac{c\alpha}{\beta}\right).$$

Integrating the first (polynomial) part yields

$$\int \left(\frac{2d}{\beta}M + \frac{2c}{\beta}\right) dM = \frac{d}{\beta}M^2 + \frac{2c}{\beta}M.$$

Let $R(M) = pM + q$, where

$$p = 2\left(b - \frac{d\alpha}{\beta}\right), \quad q = 2\left(a - \frac{c\alpha}{\beta}\right).$$

Then,

$$\int \frac{pM}{\alpha + \beta M^2} dM = \frac{p}{2\beta} \ln(\alpha + \beta M^2), \quad (\text{E6})$$

$$\int \frac{q}{\alpha + \beta M^2} dM = \frac{q}{\sqrt{\alpha\beta}} \arctan\left(M\sqrt{\frac{\beta}{\alpha}}\right). \quad (\text{E7})$$

Therefore,

$$I(M) = \frac{d}{\beta}M^2 + \frac{2c}{\beta}M + \frac{b\beta - \alpha d}{\beta^2} \ln(\alpha + \beta M^2) + \frac{2(a\beta - \alpha c)}{\beta\sqrt{\alpha\beta}} \arctan\left(M\sqrt{\frac{\beta}{\alpha}}\right) + C. \quad (\text{E8})$$

2. Steady-state solution

Substituting Eq. (E8) into Eq. (E5), we get

$$\ln P(M) = \frac{d}{\beta}M^2 + \frac{2c}{\beta}M + \left(\frac{b\beta - \alpha d}{\beta^2} - 1\right) \ln(\alpha + \beta M^2) + \frac{2(a\beta - \alpha c)}{\beta\sqrt{\alpha\beta}} \arctan\left(M\sqrt{\frac{\beta}{\alpha}}\right) + C. \quad (\text{E9})$$

Exponentiating, we obtain

$$P(M) = \mathcal{N}(\alpha + \beta M^2)^{\frac{b\beta - \alpha d}{\beta^2} - 1} \exp\left[\frac{d}{\beta}M^2 + \frac{2c}{\beta}M + \frac{2(a\beta - \alpha c)}{\beta\sqrt{\alpha\beta}} \arctan\left(M\sqrt{\frac{\beta}{\alpha}}\right)\right]. \quad (\text{E10})$$

3. Real logarithmic form for $\beta < 0$

For $\beta < 0$, define $r^2 = \alpha/|\beta|$, so that

$$\alpha + \beta M^2 = |\beta|(r^2 - M^2).$$

Using the identity

$$\arctan(iz) = i \operatorname{atanh} z = \frac{i}{2} \ln\left(\frac{1+z}{1-z}\right),$$

we can rewrite the arctangent term in Eq. (E10) as a real logarithm:

$$\frac{2(a\beta - \alpha c)}{\beta\sqrt{\alpha\beta}} \arctan\left(M\sqrt{\frac{\beta}{\alpha}}\right) = \kappa \ln\left(\frac{r+M}{r-M}\right),$$

where

$$\kappa = \frac{a\beta - \alpha c}{\beta\sqrt{\alpha|\beta|}}.$$

Dropping constants absorbed into normalization, the steady-state distribution becomes

$$P(M) = \mathcal{N}(r^2 - M^2)^s \left(\frac{r+M}{r-M}\right)^\kappa \times \exp\left(\frac{d}{\beta}M^2 + \frac{2c}{\beta}M\right), \quad |M| < r, \quad (\text{E11})$$

with parameters

$$r = \sqrt{\frac{\alpha}{|\beta|}}, \quad (\text{E12})$$

$$s = \frac{-\alpha d + b\beta}{\beta^2} - 1, \quad (\text{E13})$$

$$\kappa = \frac{a\beta - \alpha c}{\beta\sqrt{\alpha|\beta|}}. \quad (\text{E14})$$

Equation (E11) is equivalent to Eq. (E10), expressed in a real form suitable for systems with multiplicative noise amplitude $D(M) = \alpha + \beta M^2$ and $\beta < 0$.

APPENDIX F: SPATIAL MODEL DESCRIPTION

The Couzin model of collective behavior [28] simulates the movement of individuals in continuous space, with discrete time updates. Individuals update their positions ($\mathbf{x}$) and directions ($\mathbf{d}$) based on the positions and directions of their

neighbors within certain interaction zones. Thus, at any time instant, the position of i th individual is updated as

$$\mathbf{x}_i(t+1) = \mathbf{x}_i(t) + s \mathbf{d}_i(t+1), \quad (\text{F1})$$

where s is the speed of individuals. The direction vector $\mathbf{d}_i(t+1)$ is determined by various movement update rules, described in detail below:

Movement update rules. An individual's direction $\mathbf{d}_i(t+1)$ is determined by considering the following rules:

(1) **Repulsion.** If any individual is within the repulsion zone (ZOR) of a given focal individual i , the individual's next direction is given by

$$\mathbf{d}_i^{(\text{rep})}(t+1) = - \sum_{j \in \mathcal{N}_i^{(R_r)}} \frac{\mathbf{x}_j(t) - \mathbf{x}_i(t)}{\|\mathbf{x}_j(t) - \mathbf{x}_i(t)\|}.$$

Explanation:

(a) $\mathbf{d}_i^{(\text{rep})}(t+1)$: Direction vector for repulsion of individual i at time $t+1$.

(b) $\mathcal{N}_i^{(R_r)}$: Set of neighbors within the repulsion radius R_r , also called the zone of repulsion.

(c) $\mathbf{x}_j(t) - \mathbf{x}_i(t)$: Vector from the focal individual i to neighbor j .

(d) The negative sign ensures movement away from nearby individuals to avoid collisions.

(e) The individual prioritizes repulsion over all other rules, if any individual is present within this zone.

(2) **Orientation.** Individuals tend to align with the direction of neighbors within the zone of orientation zone (Ω_O) of a given focal individual i , given by the direction vector:

$$\mathbf{d}_i^{(\text{ori})}(t+1) = \sum_{j \in \mathcal{N}_i^{(R_o)}} \mathbf{d}_j(t). \quad (\text{F2})$$

Explanation:

$$\mathbf{d}_i(t+1) = \begin{cases} \mathbf{d}_i^{(\text{rep})}(t+1), \\ \mathbf{d}_i^{(\text{ori})}(t+1), \\ \mathbf{d}_i^{(\text{att})}(t+1), \\ \mathbf{d}_i^{(\text{ori})}(t+1) + \mathbf{d}_i^{(\text{att})}(t+1), \end{cases}$$

After calculating the direction considering the neighbors in different zones, the direction is normalized,

$$\mathbf{d}_i(t+1) = \frac{\mathbf{d}_i(t+1)}{\|\mathbf{d}_i(t+1)\|}.$$

A uniform angular noise term is added to the final direction to mimic biological errors in estimation and implementation. Specifically, each individual's desired direction vector $\mathbf{d}_i(t+1)$ is perturbed by an additive stochastic component according to

$$\Theta_i(t+1) = \tan^{-1} \left(\frac{d_{iy}(t)}{d_{ix}(t)} \right) + \xi_i(t),$$

$$\xi_i(t) \sim \mathcal{U} \left[-\frac{\eta\pi}{2}, \frac{\eta\pi}{2} \right],$$

$$\mathbf{d}_i(t+1) = \begin{pmatrix} \cos \Theta_i(t+1) \\ \sin \Theta_i(t+1) \end{pmatrix}.$$

(a) $\mathbf{d}_i^{(\text{ori})}(t+1)$: Direction vector for orientation of individual i at time $t+1$.

(b) $\mathcal{N}_i^{(R_o)}$: Set of neighbors within the orientation radius $R_o > R_r$ but excluding the repulsion zone and blind angle zone (see later in this section for the definition).

(c) $\mathbf{d}_j(t)$: Heading (unit velocity) vector of neighbor j at time t .

(d) This rule aligns the focal individual's direction with that of its neighbors, promoting coordinated motion.

(3) **Attraction.** Each individual tends to be attracted toward others within the attraction zone (Ω_a) of a given focal individual i , given by the direction vector:

$$\mathbf{d}_i^{(\text{att})}(t+1) = \sum_{j \in \mathcal{N}_i^{(R_a)}} \frac{\mathbf{x}_j(t) - \mathbf{x}_i(t)}{\|\mathbf{x}_j(t) - \mathbf{x}_i(t)\|}. \quad (\text{F3})$$

Explanation:

(a) $\mathbf{d}_i^{(\text{att})}(t+1)$: Direction vector for attraction of individual i at time $t+1$.

(b) $\mathcal{N}_i^{(R_a)}$: Set of neighbors within the attraction radius (R_a), which is an area outside the repulsion and orientation zones, extending to radius R_a such that $R_r < R_o < R_a$. We further exclude the blind angle zone (see later in this section for the definition).

(c) $\mathbf{x}_j(t) - \mathbf{x}_i(t)$: Vector pointing from the focal individual i toward neighbor j .

Let $n_i^{(R)}$ denote the number of neighbors of individual i within a distance R , i.e.,

$$n_i^{(R)} = \sum_{j \neq i} \mathbb{I}(\|\mathbf{x}_j(t) - \mathbf{x}_i(t)\| < R),$$

where $\mathbb{I}(\cdot)$ is an indicator function that equals 1 when the condition is true and 0 otherwise. Then, the desired direction of the focal agent is updated as

$$\begin{aligned} &\text{if } n_i^{(R_r)} > 0, \\ &\text{if } n_i^{(R_r)} = 0, \quad n_i^{(R_o)} > 0, \quad n_i^{(R_a)} = 0, \\ &\text{if } n_i^{(R_r)} = 0, \quad n_i^{(R_o)} = 0, \quad n_i^{(R_a)} > 0, \\ &\text{if } n_i^{(R_r)} = 0, \quad n_i^{(R_o)} > 0, \quad n_i^{(R_a)} > 0. \end{aligned}$$

After noise is added, each individual's desired direction may differ substantially from its previous direction. To keep turning biologically realistic, the model limits how much an individual can turn per time step by imposing a maximum turning angle $\Theta_{\max}$ or T_c . This ensures smooth trajectories and avoids unrealistically abrupt motion changes. We implement the model under periodic boundary conditions.

1. Blind angle implementation

Each individual i was assumed to have a limited field of view $\phi_{\text{view}} = 2\pi - \text{bla}$, where bla denotes the blind angle behind the individual. The relative viewing angle ϕ_{ij} between the current heading direction $\mathbf{v}_i(t)$ and the vector to neighbor j was computed as

$$\phi_{ij} = \cos^{-1} \left(\frac{\mathbf{d}_i(t) \cdot (\mathbf{x}_j(t) - \mathbf{x}_i(t))}{\|\mathbf{d}_i(t)\| \|\mathbf{x}_j(t) - \mathbf{x}_i(t)\|} \right),$$

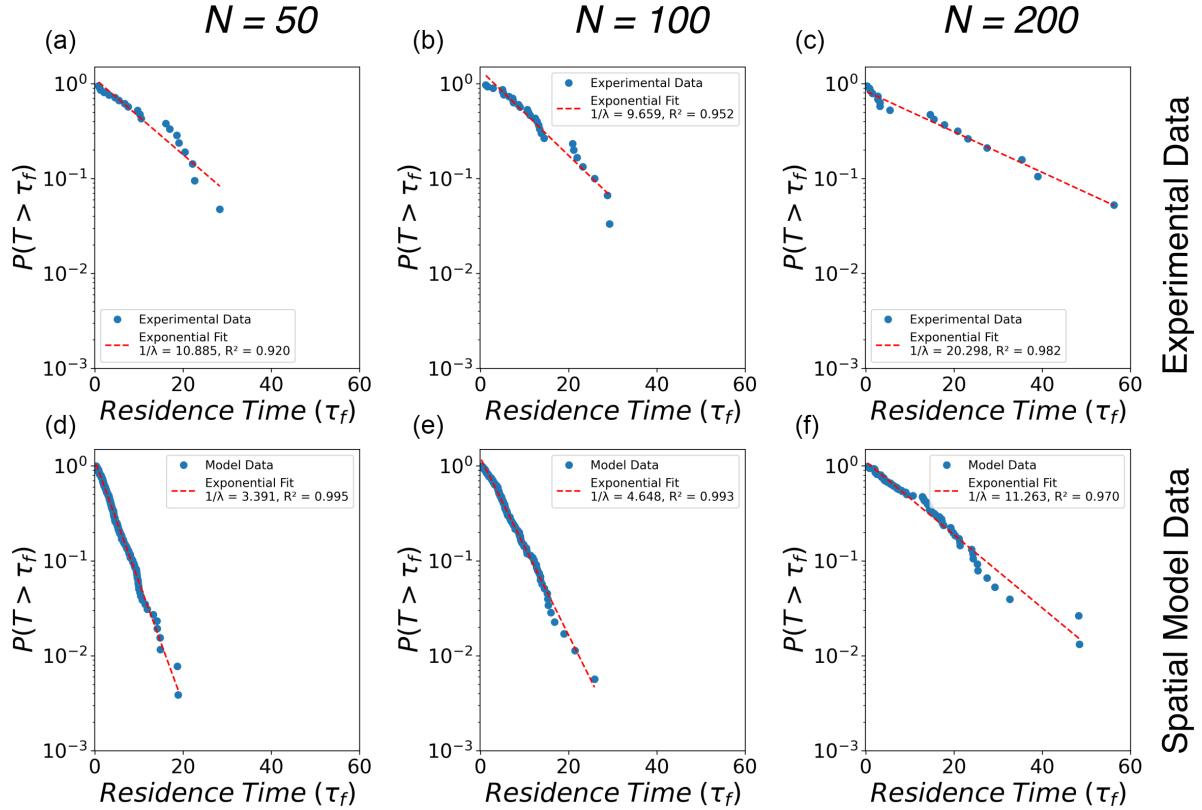


FIG. 6. Residence times from the model and the data. (a)–(c) The complementary cumulative distribution function (CCDF) of the residence time from the time series from the experimental data, and (d)–(f) for the model data for group sizes $N = 50, 100$, and 200 . The CCDF is plotted on a log scale. The dashed line is the fit to the function. Across the model and the data, a linear fit on this semilog plot indicates that the residence times are exponentially distributed. λ is the fitted exponential decay constant, and R^2 represents the goodness of the linear fits.

where $\mathbf{x}_i(t)$ and $\mathbf{x}_j(t)$ denote the positions of individuals i and j , respectively. Only neighbors within the visible field were considered in the interaction computations, i.e.,

$$|\phi_{ij}| < \pi - \frac{\text{bla}}{2}.$$

Accordingly, the attraction direction was computed as

$$\mathbf{d}_i^{\text{(att)}}(t) = \sum_{\substack{j \neq i \\ R_o \leq \|\mathbf{x}_j(t) - \mathbf{x}_i(t)\| < R_a \\ |\phi_{ij}| < \pi - \frac{\text{bla}}{2}}} \frac{\mathbf{x}_j(t) - \mathbf{x}_i(t)}{\|\mathbf{x}_j(t) - \mathbf{x}_i(t)\|}.$$

The condition $|\phi_{ij}| < \pi - \text{bla}/2$ ensures that only neighbors within the forward-facing visible field (and not behind the individual) contribute to the attraction term. The same angular filtering was also applied to the orientation interaction.

APPENDIX G: RESIDENCE TIMES

In addition to the formal SDE framework [Eq. (3)], we further compare the dynamics of the data and the model using the mean residence time or the first passage time (τ_f) of the time series of order. We determine the mean residence times numerically for both the data and the spatial model from their corresponding time series. We calculate the residence times between the two most observed states of order in the large group size ($N \leq 200$). These states are determined by

the modes of the corresponding distributions of the angular order parameter M . Figure 6 shows that the residence times distribution is exponential for both the data [panels (a)–(c)] and the model [panels (d)–(f)]. Moreover, the decay strength of the exponential distribution, denoted by λ , decreases with increasing group size (N) for both the data and the model, and consequently, the timescale of transitions ($1/\lambda$) increases with N . This quantitative comparison between the data and the model further shows that the nature of the order and the transitions observed in our data is well captured by the spatial model.

APPENDIX H: ANALYSIS OF BOUNDARY EFFECTS

As our experiments are conducted in a laboratory setup, it is essential to verify whether the experimental setup boundary affects our observed results. Previously, boundary effects have been analyzed by removing the data of individuals within a certain distance from the arena boundary [27,74]. In our case, we developed two different analyses to understand the role of the boundary, if any.

1. Cross-correlational analysis between inner and outer individuals of the swarm

We hypothesize that if the boundary has a potential role in inducing the milling behavior, individuals at the periphery

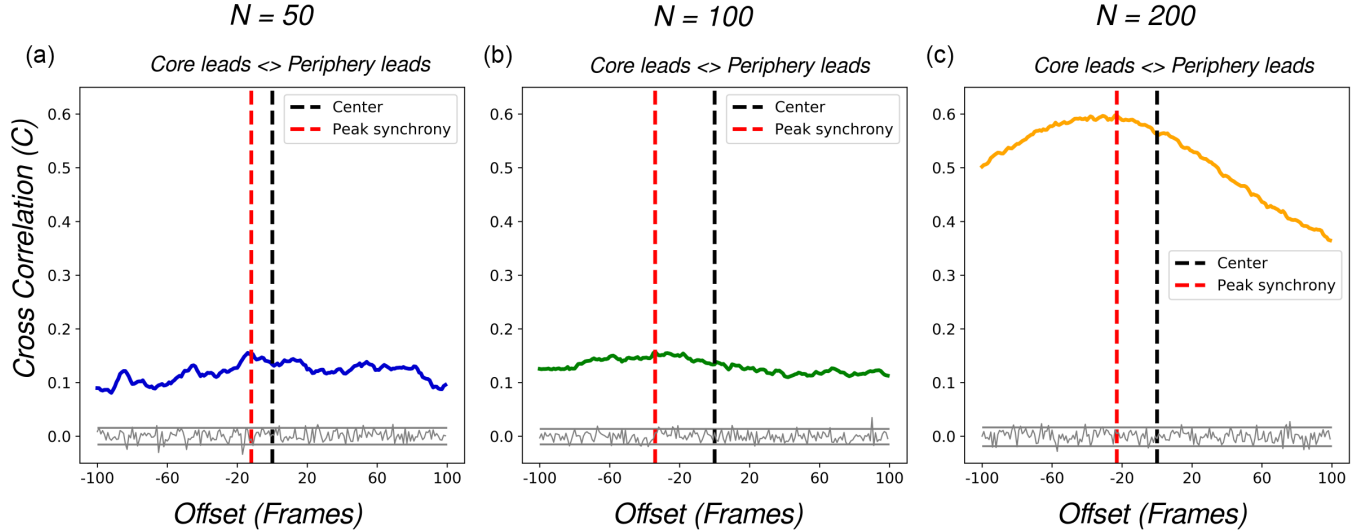


FIG. 7. A leading angular polarization for the core negates boundary influence. Cross-correlation between the angular polarization of the individuals near the center and those near the boundary as a function of time lag. A peak of this function at a negative time lag (indicated by red vertical dashed lines) strongly suggests that individuals near the center are leading. The cross-correlation is strongest in the largest group. In the smaller groups, the strength might be weak due to their already lowly polarized state. Note: 1 s = 30 frames. The gray lines are cross-correlations between two uncorrelated simulated datasets.

should start milling first, followed by individuals at the core. In other words, if we compute the milling order separately for individuals near the center and those near the periphery of the group, the two order parameters should be correlated and must be asynchronous. Specifically, if boundaries induce order, the order parameter of individuals near the periphery should precede that of individuals near the group center. To verify this, we calculated cross-correlations between the time series of angular polarization for the two subsets of individuals at different time lags. We obtained the first subset by selecting the individuals within a fixed threshold distance of 50 pixels (≈ 7 cm) from the group centroid position at any time instant, which we term the core subpopulation. Similarly, the other subpopulation, referred to as the peripheral, was selected as all individuals beyond a distance of 100 pixels (approximately 14 cm) from the group center at any given instance.

We let $M_c(t)$ and $M_p(t)$ denote the milling order parameter of the core and peripheral populations, respectively. We then compute the cross-correlation function $\mathcal{C}(\tau)$ between $M_c(t)$ and $M_p(t)$, defined as

$$\mathcal{C}(\tau) = \frac{\langle (M_c(t_i) - \mu_c)(M_p(t_i + \tau) - \mu_p) \rangle}{\sigma_c \sigma_p}, \quad (\text{H1})$$

where μ_c and μ_p are the means, while σ_c and σ_p are the standard deviations of the milling order parameters at the core and center, respectively. τ is the time offset measured as the number of time frames from the videos. According to the above equation, if cross-correlation peaks at a negative τ , then M_c leads M_p and vice versa. Figure 7 displays the cross-correlation between M_c and M_p as a function of τ for the three group sizes. In all three cases, the peak of cross-correlation, also called peak synchrony in the figure, occurs for a negative τ .

Hence, through our analysis, we find that the individuals at the center lead the individuals at the periphery in

initiating the milling, which suggests that the order arises first at the center of the group and not from the boundary (Fig. 7). Consequently, we rule out that boundaries will likely confound our results by inducing milling.

Interestingly, however, this analysis raises further questions about the mechanism of milling in beetles. For $N = 200$, M_c and M_p show a strong correlation, and the analysis suggests that individuals at the center start to rotate earlier, and the peripheral ones follow them. Since we are unable to follow a single individual for a long time, especially when they are near the center, studying this aspect further is outside the scope of our current study.

2. Spatial model with boundary

We further test for the possible effects of a boundary by implementing one in the spatial model (Fig. 8). As shown in panels (a) and (b), we implement a circular reflective boundary for two of the three group sizes ($N = 50$ and 200) studied in the data and the model. In this implementation, whenever an individual is likely to reach or cross the circular boundary, i.e., its next position's radial distance from the center of the arena is equal to or greater than the radius of the arena, then we reflect it using its current direction and the vector from its position to the center of the arena. Thus, its next direction is given by the negative of the reflection of its current direction about the radial direction from its current position:

$$d_i(t+1) = d_i(t) - 2(d_i(t) \cdot \hat{r})\hat{r}, \quad (\text{H2})$$

where $\hat{r}$ is the normalized position vector of the individual i from the center of the arena.

In Figs. 8(c) and 8(d), we show the probability distributions of M for different radii of the arena in the simulations. We hypothesize that if boundaries play a significant role in inducing milling, then reducing the size of the arena should lead to a

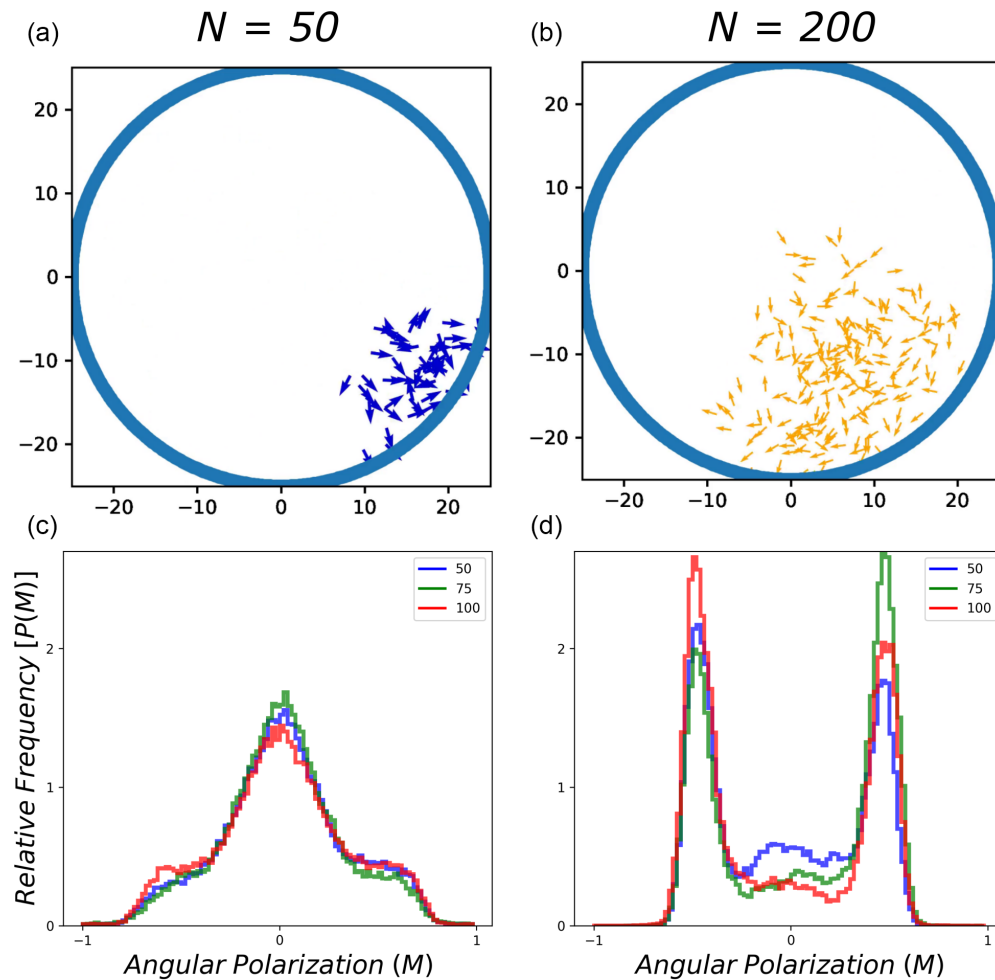


FIG. 8. Spatial model with circular boundary. (a), (b) Snapshots from the simulations of a spatial model with a circular boundary for the three group sizes considered in the study ($N = 50$ and 200). (c), (d) Probability density distributions for the two group sizes. Blue, green, and red represent an arena with a radius of 50, 75, and 100 units, respectively.

more milling order. We use three radii of size 50, 75, and 100 units to study boundary effects. We keep the other parameters the same as previously for the periodic boundary situation (Sec. V). For $N = 50$, overall, we do not observe any substantial effect of boundary on the dynamics. The group continues to show disorder for all the different radii [Fig. 8(c)]. Thus, for $N = 50$, which has disorder, reducing the arena radius does not produce order. For the larger group size, i.e., $N = 200$, from the PDF distributions, we observe an increase in the

peak height near $M = 0$ for smaller arenas. On the other hand, the peaks near the ordered states are slightly lower [Fig. 8(d)] compared to those in the larger arenas. Hence, we infer that boundaries may only decrease the order of a system and are unlikely to induce any.

Taken together, using our two analyses of boundary effects, we infer that the milling order in whirligig beetles is likely a fundamental property of intrinsic interactions among individuals, rather than an artifact of the boundary.

-
- [1] D. J. Watts and S. H. Strogatz, Collective dynamics of ‘small-world’ networks, *Nature (London)* **393**, 440 (1998).
 - [2] S. J. Altschuler, S. B. Angenent, Y. Wang, and L. F. Wu, On the spontaneous emergence of cell polarity, *Nature (London)* **454**, 886 (2008).
 - [3] F. Larroya and T. Galla, Demographic noise in complex ecological communities, *J. Phys.: Complexity* **4**, 025012 (2023).
 - [4] A. Okubo, Dynamical aspects of animal grouping: Swarms, schools, flocks, and herds, *Adv. Biophys.* **22**, 1 (1986).
 - [5] B. Partridge, The effect of school size on the structure and dynamics of minnow schools, *Anim. Behav.* **28**, 68 (1980).
 - [6] C. Becco, N. Vandewalle, J. Delcourt, and P. Poncin, Experimental evidences of a structural and dynamical transition in fish school, *Physica A* **367**, 487 (2006).
 - [7] O. Feinerman, I. Pinkoviezky, A. Gelblum, E. Fonio, and N. S. Gov, The physics of cooperative transport in groups of ants, *Nat. Phys.* **14**, 683 (2018).
 - [8] A. Kolpas, J. Moehlis, and I. G. Kevrekidis, Coarse-grained analysis of stochasticity-induced switching between

- collective motion states, *Proc. Natl. Acad. Sci. USA* **104**, 5931 (2007).
- [9] C. A. Yates, R. Erban, C. Escudero, I. D. Couzin, J. Buhl, I. G. Kevrekidis, P. K. Maini, and D. J. Sumpter, Inherent noise can facilitate coherence in collective swarm motion, *Proc. Natl. Acad. Sci. USA* **106**, 5464 (2009).
- [10] A. M. Reynolds, Intrinsic stochasticity and the emergence of collective behaviours in insect swarms, *Eur. Phys. J. E* **44**, 1 (2021).
- [11] J. Jhawar, R. G. Morris, and V. Guttal, Deriving mesoscopic models of collective behavior for finite populations, in *Integrated Population Biology and Modeling, Part B, Handbook of Statistics*, edited by S. R. Arni, S. Rao, and C. R. Rao (Elsevier, Amsterdam, Netherlands, 2019), Vol. 40, Chap. 13, pp. 551–594.
- [12] N. Miller and R. Gerlai, From schooling to shoaling: Patterns of collective motion in zebrafish (*Danio rerio*), *PLoS One* **7**, e48865 (2012).
- [13] D. Sankey, L. O'Bryan, S. Garnier, G. Cowlshaw, P. Hopkins, M. Holton, I. Fürstbauer, and A. King, Consensus of travel direction is achieved by simple copying, not voting, in free-ranging goats, *R. Soc. Open Sci.* **8**, 201128 (2021).
- [14] A. Ordemann, Vortex-swarmling of the zooplankton *Daphnia*, *Biol. Phys.* **2**, 5 (2002).
- [15] R. Mach and F. Schweitzer, Modeling vortex swarming in *Daphnia*, *Bull. Math. Biol.* **69**, 539 (2007).
- [16] S. Bazazi, K. S. Pfennig, N. O. Handegard, and I. D. Couzin, Vortex formation and foraging in polyphenic spadefoot toad tadpoles, *Behav. Ecol. Sociobiol.* **66**, 879 (2012).
- [17] P. Patra, K. Beyer, A. Jaiswal, A. Battista, K. Rohr, F. Frischknecht, and U. S. Schwarz, Collective migration reveals mechanical flexibility of malaria parasites, *Nat. Phys.* **18**, 586 (2022).
- [18] C. Castellano, S. Fortunato, and V. Loreto, Statistical physics of social dynamics, *Rev. Mod. Phys.* **81**, 591 (2009).
- [19] T. Biancalani, L. Dyson, and A. J. McKane, Noise-induced bistable states and their mean switching time in foraging colonies, *Phys. Rev. Lett.* **112**, 038101 (2014).
- [20] J. Toner and Y. Tu, Long-range order in a two-dimensional dynamical XY model: How birds fly together, *Phys. Rev. Lett.* **75**, 4326 (1995).
- [21] A. J. McKane and T. J. Newman, Stochastic models in population biology and their deterministic analogs, *Phys. Rev. E* **70**, 041902 (2004).
- [22] S. Ramaswamy, The mechanics and statistics of active matter, *Annu. Rev. Condens. Matter Phys.* **1**, 323 (2010).
- [23] S. H. Rudy, S. L. Brunton, J. L. Proctor, and J. N. Kutz, Data-driven discovery of partial differential equations, *Sci. Adv.* **3**, e1602614 (2017).
- [24] A. Nabeel, V. Jadhav, D. Raj, C. Sire, G. Theraulaz, R. Escobedo, S. K. Iyer, and V. Guttal, Data-driven discovery of stochastic dynamical equations of collective motion, *Phys. Biol.* **20**, 056003 (2023).
- [25] A. Nabeel, A. Karichannavar, S. Palathingal, J. Jhawar, D. B. Brückner, D. Raj M., and V. Guttal, Discovering stochastic dynamical equations from ecological time series data, *Am. Nat.* **205**, E100 (2025).
- [26] D. B. Brückner, A. Fink, C. Schreiber, P. J. Röttgermann, J. O. Rädler, and C. P. Brodersz, Stochastic nonlinear dynamics of confined cell migration in two-state systems, *Nat. Phys.* **15**, 595 (2019).
- [27] J. Jhawar, R. G. Morris, U. Amith-Kumar, M. D. Raj, T. Rogers, H. Rajendran, and V. Guttal, Noise-induced schooling of fish, *Nat. Phys.* **16**, 488 (2020).
- [28] I. D. Couzin, J. Krause, R. James, G. D. Ruxton, and N. R. Franks, Collective memory and spatial sorting in animal groups, *J. Theor. Biol.* **218**, 1 (2002).
- [29] H. L. Devereux, C. R. Twomey, M. S. Turner, and S. Thutupalli, Whirligig beetles as corralled active Brownian particles, *J. R. Soc. Interface* **18**, 20210114 (2021).
- [30] See Supplemental Material at <http://link.aps.org/supplemental/10.1103/kmbg-sqqc> for movies, data files, and analysis code.
- [31] MATLAB, version 9.0.0.341360 (R2016a), The MathWorks Inc., Natick, MA, 2016.
- [32] V. Guttal and C. Jayaprakash, Impact of noise on bistable ecological systems, *Ecol. Modell.* **201**, 420 (2007).
- [33] C. Gardiner, *Stochastic Methods* (Springer, Berlin, 2009), Vol. 4.
- [34] R. Friedrich, J. Peinke, M. Sahimi, and M. R. R. Tabar, Approaching complexity by stochastic methods: From biological systems to turbulence, *Phys. Rep.* **506**, 87 (2011).
- [35] L. Boninsegna, F. Nüske, and C. Clementi, Sparse learning of stochastic dynamical equations, *J. Chem. Phys.* **148**, 241723 (2018).
- [36] J. Jhawar and V. Guttal, Noise-induced effects in collective dynamics and inferring local interactions from data, *Philos. Trans. R. Soc. B* **375**, 20190381 (2020).
- [37] K. Tunström, Y. Katz, C. C. Ioannou, C. Huepe, M. J. Lutz, and I. D. Couzin, Collective states, multistability and transitional behavior in schooling fish, *PLoS Comput. Biol.* **9**, e1002915 (2013).
- [38] A. Rabani, G. Ariel, and A. Be'er, Collective motion of spherical bacteria, *PLoS One* **8**, e83760 (2013).
- [39] N. T. Ouellette, Empirical questions for collective-behaviour modelling, *Pramana* **84**, 353 (2015).
- [40] P. Bera, A. Wasim, J. Mondal, and P. Ghosh, Mechanistic underpinning of cell aspect ratio-dependent emergent collective motions in swarming bacteria, *Soft Matter* **17**, 7322 (2021).
- [41] D. Martin, H. Chaté, C. Nardini, A. Solon, J. Tailleur, and F. Van Wijland, Fluctuation-induced phase separation in metric and topological models of collective motion, *Phys. Rev. Lett.* **126**, 148001 (2021).
- [42] I. D. Couzin, J. Krause, N. R. Franks, and S. A. Levin, Effective leadership and decision-making in animal groups on the move, *Nature (London)* **433**, 513 (2005).
- [43] B. Yuce, E. Mastrocinque, M. S. Packianather, A. Lambiase, and D. T. Pham, The bees algorithm and its applications, in *Handbook of Research on Artificial Intelligence techniques and Algorithms*, edited by P. Vasant (IGI Global, 2015), pp. 122–151.
- [44] D. J. Sumpter and M. Beekman, From nonlinearity to optimality: Pheromone trail foraging by ants, *Anim. Behav.* **66**, 273 (2003).
- [45] L. Dyson, C. A. Yates, J. Buhl, and A. J. McKane, Onset of collective motion in locusts is captured by a minimal model, *Phys. Rev. E* **92**, 052708 (2015).
- [46] J. E. Herbert-Read, A. Perna, R. P. Mann, T. M. Schaefer, D. J. Sumpter, and A. J. Ward, Inferring the rules of

- interaction of shoaling fish, *Proc. Natl. Acad. Sci. USA* **108**, 18726 (2011).
- [47] D. S. Calovi, A. Litchinko, V. Lecheval, U. Lopez, A. Pérez Escudero, H. Chaté, C. Sire, and G. Theraulaz, Disentangling and modeling interactions in fish with burst-and-coast swimming reveal distinct alignment and attraction behaviors, *PLoS Comput. Biol.* **14**, e1005933 (2018).
- [48] L. Jiang, Y. Cheng, S. Gao, Y. Zhong, C. Ma, T. Wang, and Y. Zhu, Emergence of social cluster by collective pairwise encounters in *Drosophila*, *eLife* **9**, e51921 (2020).
- [49] J. G. Puckett, R. Ni, and N. T. Ouellette, Time-frequency analysis reveals pairwise interactions in insect swarms, *Phys. Rev. Lett.* **114**, 258103 (2015).
- [50] Y. Katz, K. Tunström, C. C. Ioannou, C. Huepe, and I. D. Couzin, Inferring the structure and dynamics of interactions in schooling fish, *Proc. Natl. Acad. Sci. USA* **108**, 18720 (2011).
- [51] R. Lukeman, Y.-X. Li, and L. Edelstein-Keshet, Inferring individual rules from collective behavior, *Proc. Natl. Acad. Sci. USA* **107**, 12576 (2010).
- [52] D. S. Calovi, U. Lopez, S. Ngo, C. Sire, H. Chaté, and G. Theraulaz, Swarming, schooling, milling: Phase diagram of a data-driven fish school model, *New J. Phys.* **16**, 015026 (2014).
- [53] L. Jiang, L. Giuggioli, A. Perna, R. Escobedo, V. Lecheval, C. Sire, Z. Han, and G. Theraulaz, Identifying influential neighbors in animal flocking., *PLoS Comput. Biol.* **13**, e1005822 (2017).
- [54] W. Horsthemke and R. Lefever, *Noise-Induced Transitions: Theory and Applications in Physics, Chemistry and Biology* (Springer, Berlin, 1984).
- [55] E. Dogan and E. J. Allen, Derivation of stochastic partial differential equations for reaction-diffusion processes, *Stochastic Anal. Appl.* **29**, 424 (2011).
- [56] J. Buhl, D. J. Sumpter, I. D. Couzin, J. J. Hale, E. Despland, E. R. Miller, and S. J. Simpson, From disorder to order in marching locusts, *Science* **312**, 1402 (2006).
- [57] A. Attanasi *et al.*, Information transfer and behavioural inertia in starling flocks, *Nat. Phys.* **10**, 691 (2014).
- [58] F. Ginelli, The physics of the Vicsek model, *Eur. Phys. J.: Spec. Top.* **225**, 2099 (2016).
- [59] W. L. Romey and A. R. Lamb, Flash expansion threshold in whirlingig swarms, *PLoS One* **10**, e0136467 (2015).
- [60] W. Romey, A. Walston, and P. Watt, Do 3-D predators attack the margins of 2-D selfish herds? *Behav. Ecol.* **19**, 74 (2008).
- [61] J. Jhawar, Beetles data analysis: Code, data, and movies for “Stochastic dynamics of order and disorder in milling whirlingig beetles”, Zenodo (2025), <https://doi.org/10.5281/zenodo.17645611>
- [62] H. Musoff and P. Zarchan, *Fundamentals of Kalman Filtering: A Practical Approach*, edited by P. Zarchan, H. Musoff, and F. K. Lu (American Institute of Aeronautics and Astronautics, Inc., Virginia, 2013).
- [63] C. Vamoş and M. Crăciun, Separation of components from a scale mixture of Gaussian white noises, *Phys. Rev. E* **81**, 051125 (2010).
- [64] I. Gómez *et al.*, Detecting events in time series using sliding noise tests: Application to atmospheric observations, *J. Atmos. Sci.* **71**, 1133 (2014).
- [65] Z. Tang, Y. Liu, and W. Zhang, Multi-level local mean decomposition for noise reduction in engineering signals, *Eng. Rep.* **5**, e12677 (2023).
- [66] F. J. Massey Jr, The Kolmogorov-Smirnov test for goodness of fit, *J. Am. Stat. Assoc.* **46**, 68 (1951).
- [67] M. Nagy, Z. Ákos, D. Biro, and T. Vicsek, Hierarchical group dynamics in pigeon flocks, *Nature (London)* **464**, 890 (2010).
- [68] S. M. Buchanan, J. S. Kain, and B. L. De Bivort, Neuronal control of locomotor handedness in *Drosophila*, *Proc. Natl. Acad. Sci. USA* **112**, 6700 (2015).
- [69] T. W. Dunn, Y. Mu, S. Narayan, O. Randlett, E. A. Naumann, C.-T. Yang, A. F. Schier, J. Freeman, F. Engert, and M. B. Ahrens, Brain-wide mapping of neural activity controlling zebrafish exploratory locomotion, *eLife* **5**, e12741 (2016).
- [70] T. R. Stickland, N. F. Britton, and N. R. Franks, Models of information flow in ant foraging: The benefits of both attractive and repulsive signals, in *Information Processing in Social Insects*, edited by C. Detrain, J. L. Deneubourg, and J. M. Pasteels (Birkhäuser, Basel, 1999), pp. 83–100.
- [71] A. Reina and J. A. Marshall, Negative feedback may suppress variation to improve collective foraging performance, *PLoS Comput. Biol.* **18**, e1010090 (2022).
- [72] D. T. Gillespie, The chemical Langevin equation, *J. Chem. Phys.* **113**, 297 (2000).
- [73] V. Guttal and C. Jayaprakash, Changing skewness: An early warning signal of regime shifts in ecosystems, *Ecol. Lett.* **11**, 450 (2008).
- [74] B. Lafoux, P. Bernard, B. Thiria, and R. Godoy-Diana, Confinement-driven state transition and bistability in schooling fish, *Phys. Rev. E* **110**, 034613 (2024).