

Emergent bifurcation and robust decision-making in neurodynamic swarms

Peng Ren^{1,2}, Jiang Zhao³, Pei Chi^{1,*} and Yingxun Wang¹

¹*Institute of Unmanned System, Beihang University, Beijing, China*

²*Zhongguancun Academy, Beijing, China*

³*School of Automation Science and Electrical Engineering, Beihang University, Beijing, China*



(Received 30 September 2025; accepted 7 December 2025; published 5 January 2026)

Collective motion is typically modeled by local velocity-alignment rules. However, these models do not fully capture recent evidence for complex decision-making in biological swarms and are fragile to structured, adversarial perturbations. Here, we introduce a neurodynamic swarm model where collective behavior emerges not from alignment, but from nonlinear competition within Hopfield-type ring attractors. We show that this mechanism naturally gives rise to emergent decision bifurcations that reproduce sophisticated animal behaviors. Crucially, the inherent stability of the attractors provides the swarm with remarkable resilience, allowing it to maintain coherence against adversarial perturbations that shatter alignment-based models.

DOI: [10.1103/xwdd-hjhn](https://doi.org/10.1103/xwdd-hjhn)

I. INTRODUCTION

The emergence of large-scale coordination from local interactions is a unifying principle in complex systems, underlying phenomena from cellular migration to bird flocking [1–6]. For decades, theoretical studies of such collective behavior have relied on self-propelled-particle (SPP) models, which are idealized systems in which each individual is represented as a pointlike particle that moves at a constant speed while updating its heading according to simple local rules, pioneered by Vicsek and extended by Couzin and co-workers [7–9]. In the Vicsek model, particles align their heading with the average orientation of neighbors within a metric interaction radius, up to angular noise, and thereby undergo an order-disorder transition as density or noise is varied [10]. Couzin and co-workers generalized this framework to multi-zone models with short-range repulsion, mid-range alignment, and long-range attraction, in which the effective interaction depends on distance as well as relative heading [11–14]. In addition to these canonical formulations, a broad family of physics-based agent-based flocking models has been developed, including attraction-repulsion variants, metricfree and topological interaction rules, and vision-cone-based models, which have been extensively analyzed in terms of their phase behavior and ordering transitions [15–18].

However, even for desert locust swarms, long regarded as the canonical example of an SPP system, recent decisive experiments have posed a serious challenge to this core assumption [19–23]. Observations reveal that the motion of a

single locust is not driven by simple velocity alignment, but instead relies on a more complex internal cognitive decision process [24]. This finding exposes a fundamental gap between classical models and biological reality, and compels a reexamination of the physical mechanisms underlying collective behavior.

Crucially, alignment-based (velocity-alignment) models can lose coherence under modest stochastic disturbances and are vulnerable to adversarial perturbations, here defined as structured, targeted inputs from a subset of agents that misreport or bias their states to degrade group performance [25–27]. Similar vulnerabilities are observed in both biological and engineered systems, motivating research into collective resilience mechanisms that enhance the robustness of collective behavior beyond simple neighbor-averaging rules [28–33].

To address the dual challenges of biological realism and adversarial robustness, we develop a deterministic framework based on neural ring-attractor dynamics, in which each agent's heading is represented by a localized bump of activity that moves continuously on a circular neural manifold. Unlike the probabilistic Glauber (stochastic spin-flip) dynamics used by Sayin, *et al.* to fit locust trajectories [24], our model supplies a concrete dynamical substrate. Each agent's heading is encoded in a Hopfield-type network with an explicit energy function, i.e., a symmetric recurrent circuit endowed with an energy function that ensures convergence to attractor states representing candidate headings [34]. The attractor architecture supports pattern completion and error correction, allowing the swarm to converge on coherent decisions while filtering noisy or misleading inputs [35,36]. Large-scale simulations show that, in contrast to alignment-based models that readily collapse under modest noise or adversarial perturbations, the attractor-driven swarm sustains robust global order and reproduces the decision bifurcations observed in locust experiments. These results establish ring-attractor dynamics as a biologically plausible, physically grounded, and robust substrate for collective decision-making.

*Contact author: peichi@buaa.edu.cn

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/) license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

II. NEURODYNAMIC SWARM MODEL

We model each agent's decision process by a Hopfield-type recurrent network in which the heading is encoded as a stable attractor state. The continuous-time dynamics for neuron n of agent i is given by

$$\frac{dx_{i,n}}{dt} = -x_{i,n} + \sum_{m=1}^{N_f} T_{nm} f(x_{i,m}) + U_{i,n}, \quad (1)$$

where $x_{i,n}$ denotes the membrane potential of the neuron n in agent i , T_{nm} is the synaptic weight from neuron m to neuron n within the same agent, $f(x_{i,m}) = \tanh(x_{i,m})$ is the activation function, and $U_{i,n}$ denotes the external input to neuron n in agent i .

For each agent i , we consider a Hopfield-type network with N_f units. The recurrent couplings are collected in the symmetric weight matrix $T = [T_{nm}]_{n,m=1}^{N_f}$ and $T_{nm} = T_{mn}$. For symmetric T , one can define the energy function for agent i as

$$E_i = -\frac{1}{2} \sum_{n,m=1}^{N_f} T_{nm} f(x_{i,n}) f(x_{i,m}) - \sum_{n=1}^{N_f} f(x_{i,n}) U_{i,n} + \sum_{n=1}^{N_f} \int_{f(x_{i,n})}^{f(x_{i,n})} f^{-1}(z) dz. \quad (2)$$

As shown in Appendix D, this energy satisfies $\frac{dE_i}{dt} \leq 0$, and hence the network reliably converges to a stable equilibrium for any symmetric weight matrix T . A generic symmetric matrix T typically yields a set of discrete, pointlike attractors, which are well suited for classification or pattern-completion tasks. In our setting, however, the heading is a continuous angular variable, so we require an effectively one-dimensional family of stable states rather than a small set of isolated fixed points.

To approximate such a continuum, we discretize directional space into N_f units arranged on a circle in preferred-direction space, each tuned to a preferred heading. Neighbors on the ring are coupled by local excitation and longer-range inhibition. This coupling pattern consolidates distributed sensory inputs into a single-stable bump of activity on the ring. The position of the bump encodes the chosen direction and thus implements a collective neural consensus (see Fig. 1). Although the network has a finite number of units, the bump can shift smoothly across neighboring neurons, yielding an effectively continuous representation of orientation.

Each neuron n is assigned the preferred heading

$$\alpha_n = -\pi + \frac{2\pi(n-1)}{N_f}, \quad n = 1, 2, \dots, N_f. \quad (3)$$

The symmetric weights between neurons n and m is then defined as

$$T_{nm} = \cos \left[\pi \left(\frac{\alpha_n - \alpha_m}{\pi} \right)^v \right], \quad (4)$$

with ring sharpness $v \in (0, 1]$ and $T_{nn} = 0$. By construction $T_{nm} = T_{mn}$, ensuring symmetry of the weight matrix, which is required for a well-defined Hopfield energy function. Positive couplings ($T_{nm} > 0$) mediate local excitation, whereas

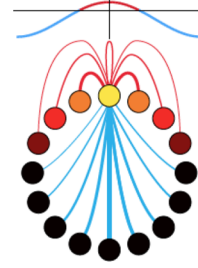


FIG. 1. Ring-attractor decision circuit. Neurons are arranged on a ring according to their preferred heading direction α_n . The yellow neuron and its neighbors form the active bump of activity encoding the current chosen direction. Red arcs indicate local recurrent excitation between neurons with similar preferred directions ($T_{nm} > 0$), whereas blue lines indicate long-range inhibition from the bump to more distant neurons (shown in black, $T_{nm} < 0$). The curve on top illustrates the cosine-shaped recurrent weight profile T_{nm} as a function of the angular difference $|\alpha_n - \alpha_m|$, which stabilizes a single bump and endows the network with a continuous attractor over heading angle.

negative couplings implement long-range inhibition. The exponent v acts as a shape parameter that controls the tuning width of this excitation-inhibition profile without changing its symmetry or peak location. For $v = 1$, Eq. (4) reduces to the cosine connectivity used in the original locust ring-attractor model [37]. Smaller values of v broaden the effective excitatory region and extend the range of distal inhibition, whereas larger values sharpen the central lobe and suppress long-range interactions. A detailed analysis of how v shapes the recurrent profile and affects the swarm dynamics is given in Sec. III (“ring sharpness”) and Appendix F.

The external input U_i encodes information from all neighbors of agent i . Each agent receives inputs from neighbors through the Hopfield-type neural network. The neighbor set is defined as $\mathcal{N}_i = \{j \mid \|p_j - p_i\| < R\}$, where p_i and p_j denote the positions of agent i and neighbor j , and R is the alignment range.

To construct the network input, each neighbor heading is projected onto a Gaussian tuning curve on the agent's angular grid, and the influence of each neighbor is weighted by distance. The input to neuron n of agent i is

$$U_{i,n} = \sum_{j \in \mathcal{N}_i} w_{ij} \times \eta_{j,n} = \sum_{j \in \mathcal{N}_i} w_{ij} \times e^{\left(-\frac{(\theta_n - \theta_j^{in})^2}{2\sigma_{in}^2} \right)}, \quad (5)$$

where $\eta_{j,n} = e^{\left(-\frac{(\theta_n - \theta_j^{in})^2}{2\sigma_{in}^2} \right)}$, $\theta_n = -\pi + \frac{2\pi(n-1)}{N_f} \in [-\pi, \pi)$ ($n = 1, 2, \dots, N_f$) is the heading of agent i , $\theta_j^{in} \in [-\pi, \pi)$ is the heading of neighbor j , σ_{in} is the width of the Gaussian tuning curve, controlling the directional selectivity, and w_{ij} is the weight of the neighbor j for agent i , which is described as follows:

$$w_{ij}(t) = \begin{cases} \frac{e^{\left(-\frac{d_{ij}^2(t)}{2\sigma_d^2} \right)}}{\sum_{j \in \mathcal{N}_i} e^{\left(-\frac{d_{ij}^2(t)}{2\sigma_d^2} \right)}}, & \|d_{ij}\| \leq R, \\ 0, & \text{otherwise,} \end{cases} \quad (6)$$

where $d_{ij} = p_i - p_j$, σ_d is a scaling parameter that determines the distance at which the interaction strength is maximized, and R is the maximum communication radius.

Based on Eq. (5), the compact form of input U_i is

$$U_i = \sum_{j \in \mathcal{N}_i} w_{ij} \times \eta_j = \sum_{j \in \mathcal{N}_i} w_{ij} \times e^{\left(-\frac{(\tilde{\theta} - \tilde{\theta}_j^{\text{in}})^2}{2\sigma_{\text{in}}^2}\right)}, \quad (7)$$

where $\tilde{\theta}_j^{\text{in}} = \theta_j^{\text{in}} \times e_{N_f \times 1}$, $\eta_j = \text{col}(\eta_{j,1}, \dots, \eta_{j,N_f})$, and $\tilde{\theta} = \text{col}(\theta_1, \dots, \theta_{N_f})$.

Combining Eqs. (1), (4), and (5), the neuronal dynamics can be obtained as follows:

$$\begin{aligned} \frac{dx_i}{dt} = & -x_i + \sum_{j=1}^{N_f} \cos \left[\pi \left(\frac{\alpha_n - \alpha_m}{\pi} \right)^v \right] \tanh(x_j) \\ & + \sum_{j \in \mathcal{N}_i} w_{ij} \times e^{\left(-\frac{(\theta_n - \theta_j^{\text{in}})^2}{2\sigma_{\text{in}}^2}\right)}. \end{aligned} \quad (8)$$

Based on Eq. (8), the decision-making dynamics of agent i can be written as follows:

$$\frac{d\tilde{x}_i}{dt} = -\tilde{x}_i + T \tanh(\tilde{x}_i) + \sum_{j \in \mathcal{N}_i} w_{ij} \times \eta_j, \quad (9)$$

where $\tilde{x}_i = [x_{i1} \ \dots \ x_{iN}]^T$ and

$$T = \begin{bmatrix} 0 & T_{12} & \dots & T_{1N_f} \\ T_{21} & 0 & \dots & T_{2N_f} \\ \vdots & \vdots & \ddots & \vdots \\ T_{N_f1} & T_{N_f2} & \dots & 0 \end{bmatrix}.$$

The compact form of the decision model (9) can be written as

$$\frac{d\tilde{x}}{dt} = -\tilde{x} + (I \otimes T) \tanh(\tilde{x}) + (W \otimes I) \times \eta, \quad (10)$$

where $\eta = \text{col}(\eta_1, \dots, \eta_N)$ and

$$W = \begin{bmatrix} 0 & w_{12} & \dots & w_{1N} \\ w_{21} & 0 & \dots & w_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ w_{N1} & w_{N2} & \dots & 0 \end{bmatrix}.$$

The decided orientation θ_i^{out} of agent i is obtained by identifying the index of the most active neuron $\hat{n}_i$ as follows:

$$\hat{n}_i = \underset{n=1, \dots, N_f}{\text{argmax}} \ x_{i,n}, \quad (11)$$

which is then mapped to the corresponding angle in the encoded directional space as follows:

$$\theta_i^{\text{out}} = \alpha_{\hat{n}_i}. \quad (12)$$

The stability of the collective dynamics, determined by the orientation states θ_i^{out} , is analyzed in Appendix E. Throughout the dynamical equations, all angular variables (such as θ_i and the ring-coordinate phases) are expressed in radians. For ease of interpretation, however, all quoted numerical angle values and plotted axis labels are reported in degrees.

Equations (11) and (12) implement a winner-take-all (WTA) readout: The decoded heading is given by the preferred direction of the maximally active neuron on the ring.

This strongly nonlinear decoder is chosen deliberately, since the role of the ring-attractor circuit in our framework is to support categorical decisions between competing directions rather than to perform a linear average of neighbor headings. Alternative readout schemes could be considered, but they tend to revert to an averaginglike behavior when multiple bumps coexist and thus weaken the winner-take-all character that is essential for the symmetry-breaking decisions studied here.

Having defined how each agent selects its orientation from the attractor network, we next embed these agents in a collective-motion framework. We consider a swarm of N agents moving in a two-dimensional plane. The motion of each agent i is governed by its neural decision state, which determines its heading θ_i^{out} and is subject to random fluctuations. The dynamics of the agent $i(t)$ are given by an overdamped Langevin equation:

$$\dot{r}_i(t) = v_0 \begin{bmatrix} \cos \theta_i^{\text{out}}(t) \\ \sin \theta_i^{\text{out}}(t) \end{bmatrix} + \delta_i(t), \quad (13)$$

where $r_i = [x_i, y_i]^T \in \mathbb{R}^2$ is the position of the agent i , v_0 is the constant self-propulsion speed, and $\delta_i(t)$ represents Gaussian white noise.

Compared with conventional direction-averaging models, the Hopfield attractor enables nonlinear competition between candidate headings, yielding sharper and more decisive orientation outcomes. This feature is particularly useful for modeling abrupt switching and multimodal decision landscapes. An extended hybrid version that augments this model with explicit attraction-repulsion zones is presented in Appendix B.

III. PARAMETER SENSITIVITY AND PERFORMANCE METRICS

In this section, we analyze how the dynamics depend on key model parameters. To describe the performance of the model, we define the following global measures:

Order parameter: The degree of directional alignment among all agents, i.e.,

$$P_{\text{ord}} = \frac{1}{Nv_0} \left\| \sum_{i=1}^N v_i \right\|.$$

Convergence time: The time required for swarm alignment, i.e.,

$$T_{\text{ord}} = \{t | P_{\text{ord}}(t) > 0.95\}.$$

With these performance measures in place, we now describe how the main model parameters affect the collective dynamics.

Ring sharpness v : The exponent v sets the sharpness of the excitation-inhibition profile in the ring attractor. Small v leads to global inhibition, while larger v produces localized excitation with distal inhibition (see Fig. 2). Simulations show that an intermediate range $v \in [0.3, 0.7]$ achieves fast convergence, smooth transitions, and robust alignment. Full-parameter sweeps are reported in Appendix F.

Communication radius R : The agents' interaction network is well approximated by a random geometric graph [38,39].

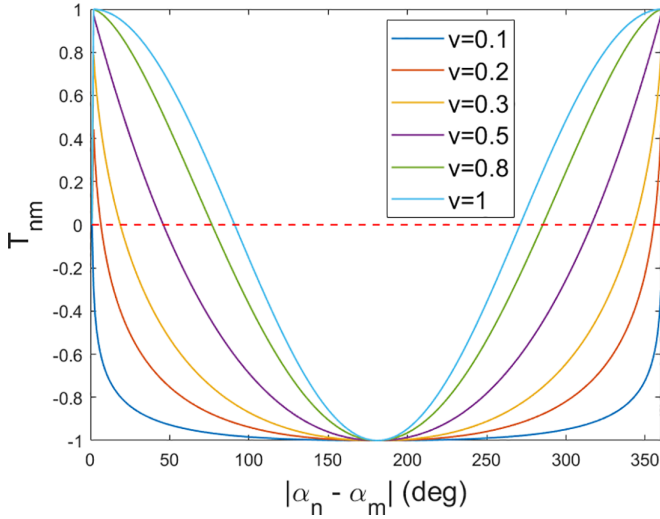


FIG. 2. Synaptic weight profiles T_{nm} for different sharpness exponents v . For $v \leq 0.1$, all connections are inhibitory ($T_{nm} < 0$). Increasing v induces a transition from global inhibition to a mixed excitation-inhibition regime. Smaller values of v yield broader coupling, whereas larger values produce sharper, more localized activity bumps. The curves are symmetric around 180° because T_{nm} depends only on the angular separation $|\alpha_n - \alpha_m|$ and is an even function of this difference [see Eq. (4)].

On an $L \times L$ domain, a classical sufficient condition for high-probability connectivity is

$$R > L \sqrt{\frac{\ln N}{\pi N}}.$$

This bound provides a conservative estimate of the minimum communication radius needed to sustain connectivity and enable global consensus. See Appendix C for a brief derivation and its extension to open-space conditions.

Input width σ_{in} : The width σ_{in} of the directional input strongly affects performance. Narrow inputs make the swarm sensitive to local fluctuations and increase directional variability, whereas very broad inputs blur directional selectivity and slow down convergence. Simulations indicate an optimal operating range $\sigma_{in} \in [50^\circ, 150^\circ]$. Parameter sweeps are reported in Appendix F.

Neighbor weighting scale σ_d : The spatial weighting scale σ_d sets the strength and range of neighbor influence and thus controls the balance between information pooling and locality. Parameter sweeps identify an optimal band $\sigma_d \in [0.6L, 0.9L]$. When σ_d is too small, neighbor information is localized and fragmented, which increases convergence time. When σ_d is too large, inputs are oversmoothed, and the activity bump broadens. The full-parameter sweeps are reported in Appendix F.

Number of neurons N_f : The parameter N_f controls the directional discretization of the ring-attractor circuit, i.e., the number of discrete heading channels available for the winner-take-all readout. To assess whether this discretization affects the macroscopic flocking behavior, we repeated the noise-robustness sweep for $N_f = 90, 180, 360$, while keeping all other parameters fixed. Figure 3 shows the stationary order parameter P_{ord} as a function of the noise amplitude δ_i for

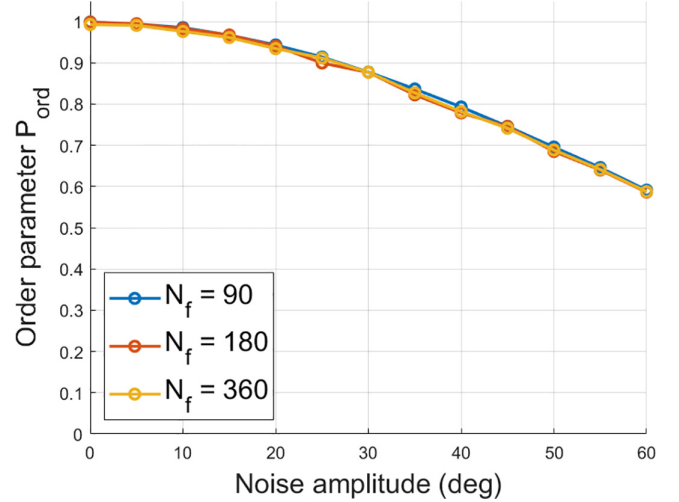


FIG. 3. Noise-robustness curves for different directional resolutions N_f . The overlap indicates that the stationary order parameter P_{ord} is essentially independent of N_f .

these three cases, with each data point averaged over 20 independent realizations and all other parameters identical to those used in other parameters sweep. The curves essentially overlap, indicating that the noise-robustness properties of the neural decision model are insensitive to N_f once it is sufficiently large. Increasing N_f mainly refines the directional resolution of the internal representation, without materially changing the macroscopic order-noise relationship.

IV. EMERGENT DECISION BIFURCATION

To validate the neurodynamic framework against behavioral data, we reproduce the paradigm of Sayin *et al.* [24], in which a focal locust is placed between two fixed targets while the lateral separation D is varied. For each D , we estimate the steady-state marginal distribution of the focal agent's lateral coordinate y from an ensemble of 200 independent trials. Figure 4 shows a behavioral transition as D is increased. For small D , the distribution is unimodal and concentrated near $y \approx 0$, consistent with an averaging strategy. For $D \geq 2.7$ cm, the central peak loses stability and two symmetric off-axis peaks emerge, consistent with a winner-take-all choice. This bifurcation reproduces the behavior reported for locusts and is emergent in the sense that the left/right choice is not prescribed by an explicit rule but arises from the interaction between the internal ring-attractor circuit and the externally controlled geometric parameter D .

Mechanistically, the transition corresponds to a bifurcation in the agent's internal energy landscape. Beyond D_c , the symmetric central attractor loses stability and a pair of asymmetric stable attractors appears. The eventual selection of one branch is driven by small perturbations such as intrinsic neural noise or variability in initial conditions, reflecting spontaneous symmetry breaking and the stochastic nature of biological decisions.

In standard velocity-averaging models, two symmetric targets generate an effective input whose mean direction remains close to the spatial midline. Thus, the focal agent

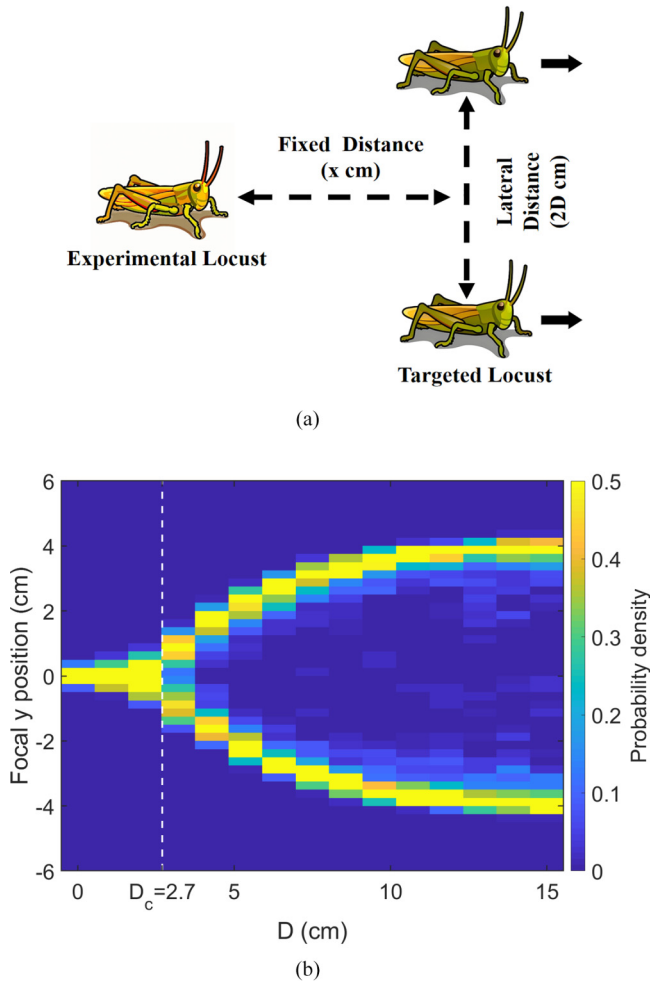


FIG. 4. Decision bifurcation in the neurodynamic model. (a) Schematic of the two-choice assay: a focal locust chooses its lateral position relative to two fixed targets separated by a controllable lateral distance D , which acts as the external control parameter. (b) Bifurcation diagram of the model. The colormap shows the steady-state probability density of the focal agent's lateral position y as a function of D (200 trials per value). The transition from a single-central band to two symmetric off-axis bands beyond $D_c \approx 2.7$ cm reflects a bifurcation in the stability of the internal attractors as D is varied.

tends to follow an averaged trajectory, and a robust bi-modal choice distribution does not emerge. Compared with probabilistic Glauber dynamics [24] that reproduce the bifurcation at the level of trajectory statistics but lack an explicit low-dimensional energy landscape, the present ring-attractor makes the underlying symmetry-breaking mechanism and the associated bifurcation structure more transparent.

V. ROBUSTNESS TO STOCHASTIC NOISE

We compare the noise robustness of the attractor-driven model to three baselines: (1) an alignment-only Vicsek model, (2) a full-interaction Couzin model (repulsion/orientation/attraction zones), and (3) a probabilistic Glauber-type rule from Ref. [24]. The precise update rules for all three models are summarized in Appendix A. To isolate

interaction effects, we simulate two variants of the neural model. The first omits short-range repulsion and long-range attraction and is used for comparisons with the Vicsek model. The second augments the neural model with Couzin-style repulsion and attraction applied to the velocity command for comparisons with the Couzin model (see Appendix B). For fair comparisons, the numerical parameter choices for all models are summarized in Appendix G. Figures 5(a)–5(c) show ensemble means over 50 realizations. Across these conditions, the attractor-driven swarm maintains a higher order parameter under noise, whereas other models lose coherence.

To quantitatively characterize the robustness of our model, we analyze the system's phase transition as a function of noise intensity. The order parameter P_{ord} is plotted as a function of the noise amplitude (in degrees) in Fig. 6. Noise is modeled as an additive perturbation to the decoded heading of each agent at every time step, drawn from a zero-mean Gaussian distribution with standard deviation a . The curve shows the ensemble mean of P_{ord} averaged over 30 independent experiments for a swarm of $N = 20$ agents, with the shaded band indicating standard error of the mean (s.e.m.). As the angular noise a increases, the system undergoes a continuous transition from a highly ordered state ($P_{\text{ord}} \approx 1$) to disorder. A robust ordered phase, defined here as $P_{\text{ord}} \geq 0.85$, persists up to $a \approx 35^\circ$; beyond this point, the collective coherence decays steadily. Over the same noise sweep, the neural decision model substantially outperforms the Vicsek, Couzin, and Glauber-based baselines, which lose global order at markedly lower noise. These results demonstrate that the neural decision model is highly robust to stochastic noise.

VI. RESILIENCE TO ADVERSARIAL PERTURBATIONS

In collective-motion studies, controlled manipulations of communicated states have long been used to probe swarm robustness, decision-making thresholds, and effective-temperature-like responses, typically through contrarian or biased agents and related fault models [25–29]. Motivated by this line of work, we examine the resilience of our model to targeted informational attacks by introducing a fraction ρ of adversarial agents that follow the nominal-motion dynamics but deliberately falsify the state information they broadcast to neighbors (e.g., reporting an incorrect heading or velocity).

Rather than adding unstructured random fluctuations, these perturbations selectively corrupt the interaction signals exchanged between agents. They directly degrade the information integrity on which the distributed decision rules rely and can be viewed as a flocking analog of Byzantine faults in distributed systems [40].

We focus on two representative adversarial perturbation protocols that are used throughout the paper:

(1) *Onset orthogonal misreporting.* A fraction $\rho = 0.30$ of agents are adversarial from $t = 0$ and consistently report headings rotated by 90° from their true directions. This protocol delivers a strong, persistent informational perturbation to the swarm. Unless stated otherwise, simulations use periodic boundaries with $N = 20$, $v_0 = 2$, $L = 10$, $r_r = 0.1$, $r_o = 2.9$, $r_a = 3.0$, and $R = 3$. A small amount of background angular noise (standard deviation 0.1°) is applied to all models.

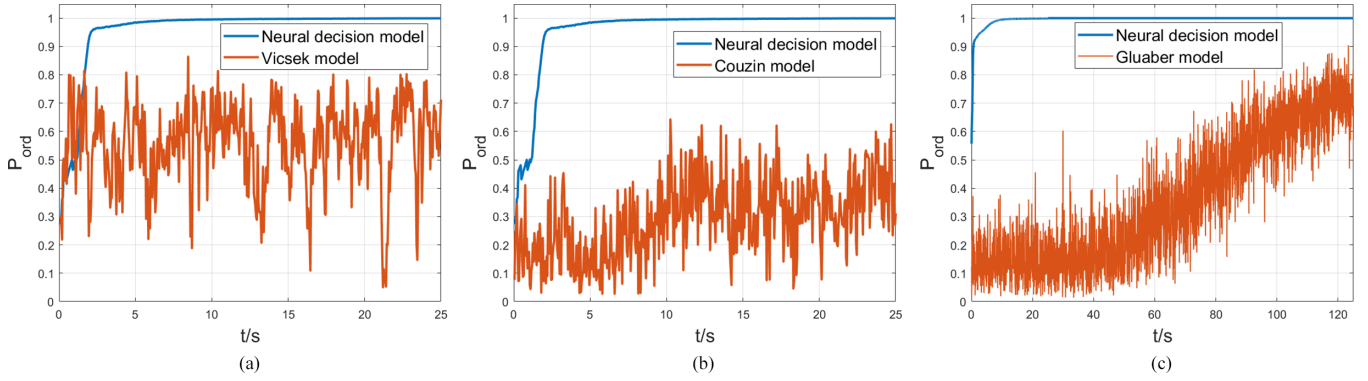


FIG. 5. Noise-robustness comparison. (a) Alignment-only baseline (Vicsek without repulsion) vs attractor-driven model (no repulsion). (b) Full-interaction comparison: Couzin model (zone radii $r_r = 0.4$, $r_o = 2.6$, $r_a = 3$) vs attractor-driven model augmented with equivalent short-range repulsion and long-range attraction. (c) Probabilistic Glauber dynamics vs attractor-driven model. Curves are ensemble means over 50 independent realizations; all simulations use identical additive noise (uniform in $[0, 0.8]$).

(2) *Postconsensus drift (stealth) perturbation.* A coherent swarm of $N = 50$ agents is first allowed to reach consensus. At time $t = t_{\text{inj}} = 20$ s, a small fraction $\rho_{\text{small}} = 0.05$ of agents switch to adversarial mode and each reports a slight, constant bias $\Delta\theta = 2^\circ$ while maintaining nominal kinematics. This protocol probes the system's ability to suppress subtle, cumulative manipulation designed to evade immediate detection with $v_0 = 2$, $L = 10$, $R = 3$, $(r_r, r_o, r_a) = (0.1, 2.9, 3.0)$, and the same background noise.

Figure 7(a) compares the three models under the onset attack. The attractor-driven swarm self-organizes into a coherent state despite the strong adversarial fraction, although the collective heading is biased by the attackers. In contrast, the Vicsek and Couzin baselines fail to establish stable consensus and instead devolve into disordered millinglike states. Figures 7(b) and 7(c) summarize the postconsensus experiment. All models maintain cohesion, with order parameter near unity, yet the attractor-driven model suppresses cumulative drift,

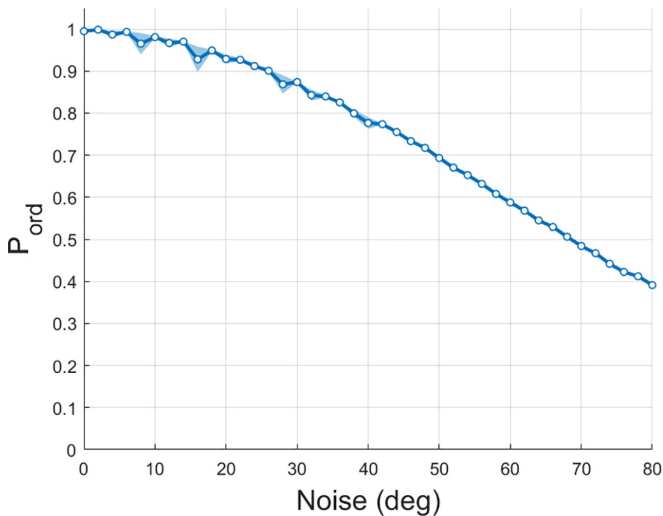


FIG. 6. Noise-order relation in the neural decision model. Stationary order parameter P_{ord} vs noise amplitude $\delta_i(t)$, illustrating the transition from a highly ordered to a disordered regime as noise increases.

keeping the final deviation below 1° . At $t = 100$ s, the Vicsek and Couzin models amplify the small bias into large runaway deviations: Under the conditions shown, the Couzin deviation reaches about 120° and the Vicsek deviation about 170° . The corresponding swarm dynamics are provided in the Supplemental Material [41].

To quantify our model's resistance to such perturbations, we examine the steady-state order parameter and collective heading deviation as functions of the adversarial fraction. The sweep results are summarized in Fig. 8.

We first probe the swarm's tolerance to adversaries present from $t = 0$. In this protocol, a fraction ρ of agents are designated adversarial and persistently report headings rotated by 90° relative to the true emergent consensus direction. Figure 8(a) shows the steady-state order parameter P_{ord} as a function of ρ . The neurodynamic swarm displays the pronounced resilience: $P_{\text{ord}} > 0.9$ up to the critical fraction $\rho_c \approx 0.35$. For $\rho > \rho_c$, the collective order declines but the decline is gradual rather than an abrupt collapse. This behavior indicates that the attractor readout substantially attenuates strongly conflicting neighbor reports during consensus formation.

We next probe a subtler, experimentally relevant threat: a postconsensus perturbation in which a coherent swarm is infiltrated by a minority of stealthy adversaries. The swarm is allowed to converge before $t = 20$ s. At $t = 20$ s, a fraction ρ of agents begins to report a small, persistent bias $\Delta\theta = 2^\circ$ while otherwise behaving normally. Figure 8(b) shows that the swarm's internal coherence, measured by the order parameter, remains effectively unchanged across the tested range of ρ , with $P_{\text{ord}} \approx 1$. In other words, the attractor readout preserves high internal alignment even when a minority continuously injects a subtle bias.

Figure 8(c) shows the steady-state heading deviation relative to the preattack baseline. We observe a nonlinear response: For small adversary fractions $\rho \leq 0.15$, the swarm effectively ignores the bias and the drift is negligible. Above this threshold, the persistent bias accumulates and the group is progressively steered away from its original course. These results reveal a trade-off. The attractor architecture robustly preserves internal cohesion through error correction at the

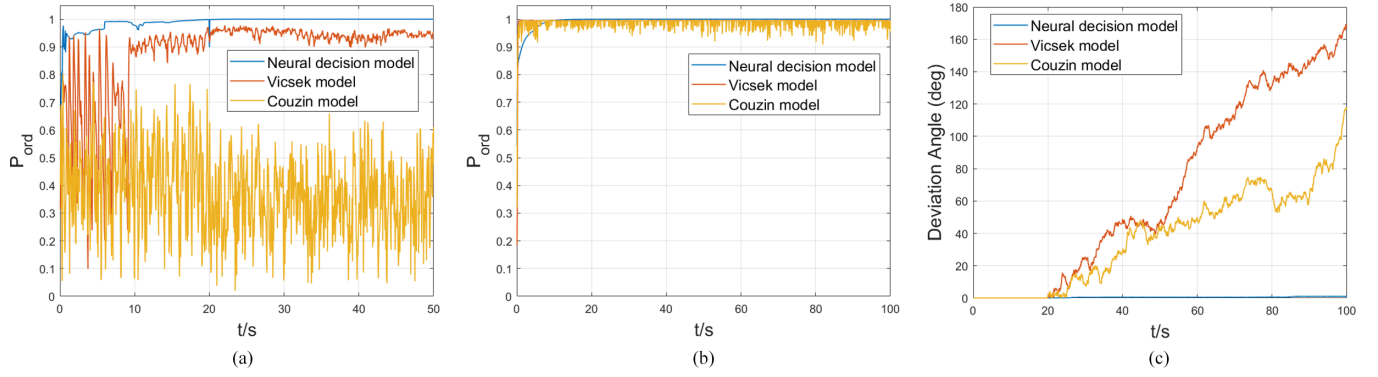


FIG. 7. Resilience to adversarial perturbations. (a) Onset attack. Adversaries comprising 30% of the population are present from $t = 0$ s and report headings rotated by 90° . Shown are ensemble means of the order parameter P_{ord} for the attractor-driven model, the Vicsek baseline, and the Couzin baseline over 50 realizations. (b) Postconsensus order. After a coherent swarm is established, 5% stealthy adversaries are injected at $t = 20$ s reporting a 2° bias. Shown is the order parameter P_{ord} . (c) Heading shift. Time evolution of the collective heading deviation, defined as the angle between the instantaneous mean heading and the preattack baseline.

readout level. However, when a coherent minority exerts a sustained bias, the collective heading can be gradually corrupted.

VII. CONCLUSIONS

We have introduced a deterministic neurodynamic framework in which each agent's heading decision emerges from a Hopfield-type ring-attractor network that implements competitive dynamics and pattern completion. Extensive numerical simulations have reproduced the decision bifurcation observed in locust experiments. Crucially, unlike conventional alignment rules, the model has maintained global coherence and suppressed bias under both stochastic noise and targeted informational perturbations. These findings suggest that ring-attractor dynamics can provide a physically realizable basis for more resilient collective decision-making and may offer useful insights for understanding biological group behavior and for designing fault tolerant multiagent systems.

DATA AVAILABILITY

No external datasets were used. The raw experimental data supporting the findings of this study are available from the corresponding author upon reasonable request.

APPENDIX A: BASELINE MODEL DEFINITIONS

1. Vicsek model

The Vicsek model [7] is described as follows:

$$\begin{aligned}\theta_i(t + \Delta t) &= \langle \theta_j(t) \rangle_{|r_j - r_i| < R} + \xi_i(t), \\ \mathbf{v}_i(t) &= \begin{bmatrix} v_0 \cos \theta_i(t) \\ v_0 \sin \theta_i(t) \end{bmatrix}, \\ \mathbf{r}_i(t + \Delta t) &= \mathbf{r}_i(t) + \mathbf{v}_i(t) \Delta t,\end{aligned}\quad (\text{A1})$$

where θ_i is the orientation angle of the i th particle, $\langle \theta_j(t) \rangle_{|r_j - r_i| < R} = \arctan\left(\frac{\sum_{j \in N_i(t)} \sin \theta_j(t)}{\sum_{j \in N_i(t)} \cos \theta_j(t)}\right)$ is the average within

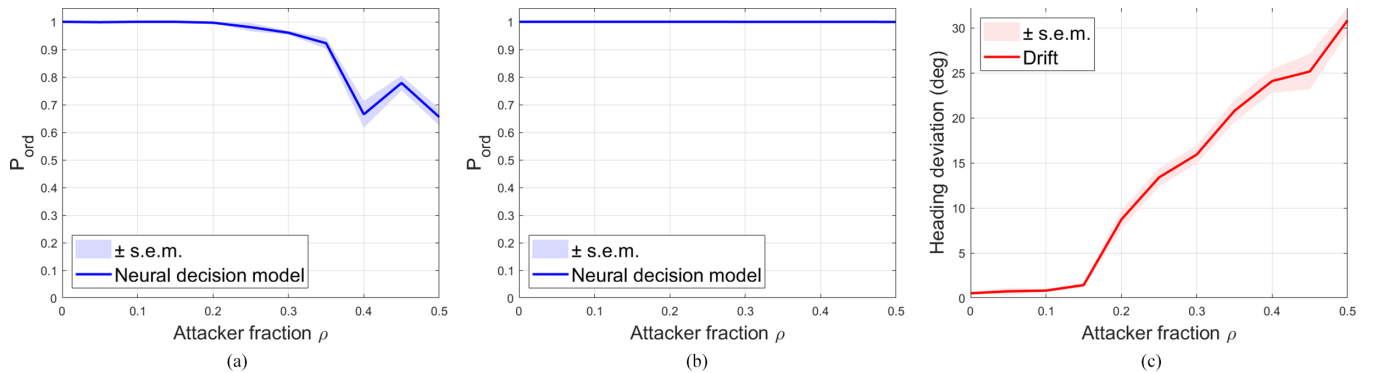


FIG. 8. Quantitative characterization of resilience to adversarial perturbations. (a) Onset perturbation. The plot shows the steady-state order parameter P_{ord} vs adversary fraction ρ when a fraction ρ of agents are adversarial from $t = 0$ and consistently report headings rotated by 90° . The neurodynamic swarm preserves high coherence ($P_{\text{ord}} \geq 0.9$) up to a critical fraction $\rho \approx 0.35$; for $\rho > \rho_c$, the order decreases continuously rather than collapsing catastrophically. (b), (c) Postconsensus (stealth) perturbation. A coherent swarm is allowed to converge; at $t = 20$ s a fraction ρ of agents begin reporting a small persistent bias $\Delta\theta = 2^\circ$. Panel (b) shows the steady-state order, which remains approximately unity for tested ρ up to 0.5. Panel (c) shows the steady-state heading deviation (mean angular difference from the preattack baseline) as a function of ρ : for $\rho \leq 0.15$, the drift is negligible, while for larger ρ the persistent bias accumulates and the group is progressively steered away from its original heading. All curves are ensemble means over 30 independent realizations; shaded bands denote $\pm$ s.e.m.

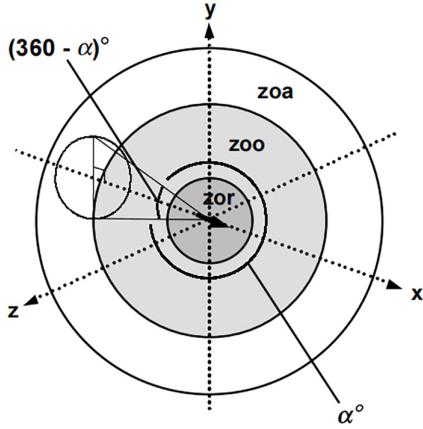


FIG. 9. A schematic of the three concentric sensory zones for an individual agent (centered at the origin): the ZOR, the zone of orientation (ZOO), and the zone of attraction (ZOA). A potential blind volume behind the agent, characterized by the angle α , is also shown.

radius R neighborhood, and uniform random noise in $[-\pi, \pi]$. v_0 is the constant speed and is the position vector of the particle.

2. Couzin model

The Couzin model [8] describes the collective behavior of a swarm through a set of biologically inspired, local interaction rules. Each agent i is governed by its interactions with neighboring agents, which are categorized into three concentric zones based on their distance from agent i , as depicted in Fig. 9. The decision-making process for agent i at each time step t follows a strict priority hierarchy. First, agent i assesses the zone of repulsion (ZOR), defined by a radius r_r . If any neighbors j are detected within this zone ($0 < \|r_{ij}(t)\| \leq r_r$), the agent's primary imperative is to move away from them, where r_{ij} is the distance between agent i and agent j . The desired direction vector due to repulsion $\mathbf{d}_r$ is computed as the sum of unit vectors pointing away from each of the n_r neighbors in the ZOR:

$$\mathbf{d}_r(t + \tau) = - \sum_{j \neq i}^{n_r} \frac{r_{ij}(t)}{\|r_{ij}(t)\|}. \quad (\text{A2})$$

If the ZOR is empty, agent i then considers neighbors within the ZOO (defined by $r_r < \|r_{ij}(t)\| \leq r_o$) and the ZOA (defined by $r_o < \|r_{ij}(t)\| \leq r_a$). The desired direction of orientation $\mathbf{d}_o$ is the average direction of motion of the n_o neighbors within the ZOO:

$$\mathbf{d}_o(t + \tau) = \sum_{j \neq i}^{n_r} \frac{v_j(t)}{\|v_j(t)\|}, \quad (\text{A3})$$

and the desired direction of attraction $\mathbf{d}_a$ is the vector pointing toward the center of mass of the n_a neighbors within the ZOA, which is computed by averaging the unit vectors:

$$\mathbf{d}_a(t + \tau) = \sum_{j \neq i}^{n_r} \frac{r_{ij}(t)}{\|r_{ij}(t)\|}, \quad (\text{A4})$$

where v_j is the speed of agent j .

If neighbors are present within the ZOR, the individual moves to avoid collisions and the direction is set as $\mathbf{d}_i(t + \tau) = \mathbf{d}_r(t + \tau)$. If no neighbors are in the ZOR but are found exclusively within the ZOO, then $\mathbf{d}_i(t + \tau) = \mathbf{d}_o(t + \tau)$. If neighbors are found exclusively within the ZOA, the direction becomes $\mathbf{d}_i(t + \tau) = \mathbf{d}_a(t + \tau)$. In cases where neighbors are simultaneously located in both the ZOA and ZOO, the desired direction is computed as the average of the attraction and orientation components: $\mathbf{d}_i(t + \tau) = \frac{1}{2}(\mathbf{d}_a(t + \tau) + \mathbf{d}_o(t + \tau))$. If no neighbors are detected in any zone, then $\mathbf{d}_i(t + \tau) = \mathbf{d}_i(t)$.

The velocity of agent i is described as follows:

$$v_i(t) = s \frac{\mathbf{d}_i(t)}{\|\mathbf{d}_i(t)\|}, \quad (\text{A5})$$

where s is the speed of the agent i .

3. Glauber-based neurodynamic swarm model

For direct comparison, we implemented the neurodynamic swarm model introduced in Ref. [24]. In this model, the decision-making process of each agent is governed by a ring attractor composed of N_s binary neurons $\sigma_n \in \{-1, +1\}$. The network state evolves to minimize the Hamiltonian

$$H = -\frac{1}{N_s} \left(\sum_{n,m} J_{nm} \sigma_n \sigma_m + \sum_n h_n \sigma_n \right), \quad (\text{A6})$$

where $J_{nm} = \cos[\pi(\frac{|\alpha_n - \alpha_m|}{\pi})^v]$, $\alpha_n = \frac{2\pi n}{N_s}$, and h_n represents the external input. When an agent perceives stimuli at angles $\{\theta_r\}$, the external field is given by a superposition of Gaussian bumps,

$$h_n = \frac{h_0}{\sqrt{2\pi}\sigma^2} e^{\frac{\alpha_j - \theta_j}{\sqrt{2\pi}\sigma^2}}. \quad (\text{A7})$$

The Glauber dynamics [42] is used to simulate the network's dynamics. The spin-flip probability is

$$p(\Delta E) = \begin{cases} 1, & \Delta E \leq 0, \\ e^{-\Delta E/T}, & \Delta E > 0, \end{cases} \quad (\text{A8})$$

where T is the temperature. This procedure is repeated until the network reaches statistical equilibrium (typically after $50N_s$ updates).

The agent's velocity is updated according to the equilibrium state of its neural network as

$$v_i = \frac{v_0}{N_s} \sum_{\text{spins}} \alpha_{i,n}, \quad (\text{A9})$$

where $\alpha_{i,n}$ denotes the preferred direction of spin n . In addition, each agent can perceive the orientation of all agents.

APPENDIX B: HYBRID ATTRACTION-REPULSION NEURODYNAMIC MODEL

To compare with the Couzin model, we add repulsion and attraction in the neurodynamics. Consider a multiagent system composed of N followers. The dynamics of follower $i \in \{1, \dots, N\}$ is

$$\dot{p}_i(t) = v_i(t), \quad v_i(t) = u_i(t), \quad (\text{B1})$$

where $p_i = [x_i, y_i]^T \in \mathbb{R}^2$ denotes the position of agent i and $v_i = [v_{ix}, v_{iy}]^T \in \mathbb{R}^2$ denotes the speed of agent i .

The control law u_i for each agent is constructed from three components.

1. Velocity alignment

The alignment behavior encourages the agent to move in the direction determined by its neural decision-making system. The velocity alignment can be written as follows:

$$v_i^{\text{ali}}(t) = v_0 \begin{bmatrix} \cos \theta_i^{\text{out}}(t) \\ \sin \theta_i^{\text{out}}(t) \end{bmatrix}, \quad (\text{B2})$$

where v_0 is a constant number.

2. Repulsion speed

To avoid collisions, a repulsive velocity is generated when neighboring agents are within a critical distance:

$$v_i^{\text{rep}}(t) = \begin{cases} K_{\text{rep}} \sum_{j \in N_{\text{nei}}} (r_r - \|d_{ij}\|) \frac{d_{ij}}{\|d_{ij}\|}, & \|d_{ij}\| < r_r, \\ 0, & \text{otherwise,} \end{cases} \quad (\text{B3})$$

where K_{rep} is the proportional gain parameter and r_r is the maximum interaction range for the repulsion.

3. Attraction speed

To avoid collisions, a repulsive velocity is generated when neighboring agents are within a critical distance,

$$v_i^{\text{att}}(t) = \begin{cases} K_{\text{att}} \sum_{j \in N_{\text{nei}}} (r_o - \|d_{ij}\|) \frac{d_{ij}}{\|d_{ij}\|}, & r_o < \|d_{ij}\| \leq r_a, \\ 0, & \text{otherwise,} \end{cases} \quad (\text{B4})$$

where K_{att} is the proportional gain parameter and r_a is the maximum interaction range for the attraction.

Combining Eqs. (B2)–(B4), the desired controller can be written as follows:

$$\tilde{u}_i(t) = v_i^{\text{ali}}(t) + v_i^{\text{rep}}(t) + v_i^{\text{att}}(t). \quad (\text{B5})$$

Based on Eq. (B5), we also introduced a cutoff at v_{max} , keeping the direction of the desired velocity but reducing its magnitude if it is over the limit. Therefore, the controller can be described as follows:

$$u_i(t) = \frac{v_i}{\|v_i\|} \min\{\tilde{u}_i, v_{\text{max}}\}. \quad (\text{B6})$$

APPENDIX C: CONNECTIVITY ANALYSIS OF THE HYBRID SWARM MODEL

The communication radius R is a critical parameter that determines the connectivity of the agent interaction network and thus the possibility of global consensus. In this Appendix, we analyze the requirements on R under open-space conditions.

While the bound in Eq. (C1) applies to periodic domains with fixed system size L , the situation is more intricate in open space. In this case, $L(t)$ is not externally prescribed but emerges as the swarm's spatial diameter, which evolves dynamically over time. Accordingly, the connectivity condition

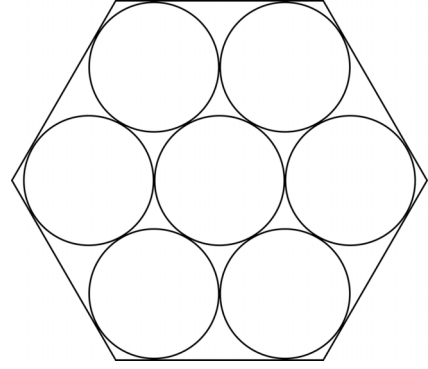


FIG. 10. The agents are arranged in the hexagonal pattern.

generalizes to

$$R(t) > L(t) \sqrt{\frac{\ln N}{N\pi}}, \quad (\text{C1})$$

where $L(t)$ denotes the swarm's spatial diameter at time t . Since $L(t)$ varies dynamically during swarm evolution, we seek an upper bound $\tilde{L}$ such that $\tilde{L} \geq L(t)$, to derive a conservative estimate for the minimal communication radius.

To determine $\tilde{L}$, we model the swarm as a planar packing of N nonoverlapping discs, each with radius r_r imposed by the short-range repulsion mechanism. The total occupied area is approximately $N\pi R_{\text{rej}}^2$, which corresponds to a square of side length as follows:

$$L^2 \approx N\pi r_r^2 \implies L \approx r_r \sqrt{\pi N}.$$

However, this assumes idealized full coverage. To account for realistic spatial gaps, we consider the hexagonal lattice, which represents the densest known two-dimensional packing and has a maximum packing fraction of $\phi_{\text{max}} \approx 0.9069$ (see Fig. 10). Accounting for this yields a tighter bound:

$$\tilde{L} \approx r_r \sqrt{\frac{\pi N}{\phi_{\text{max}}}} = r_r \sqrt{\frac{\pi N}{0.9069}}. \quad (\text{C2})$$

Substituting Eq. (C3) into Eq. (C1), we obtain the lower bound for the communication radius as follows:

$$R(t) > r_r \sqrt{\frac{\ln N}{0.9069}}. \quad (\text{C3})$$

This bound captures the expected spatial footprint of a swarm subject to local repulsion, ensuring high-probability connectivity without excessive communication range provisioning.

APPENDIX D: STABILITY ANALYSIS OF THE NEURODYNAMIC DECISION CORE

The energy function of the ring-attractor Hopfield-type dynamics can be written as follows:

$$E_i = - \sum_{n=1}^{N_f} \sum_{m=1}^{N_f} \frac{1}{2} T_{nm} f(x_{i,m}) f(x_{i,n}) - \sum_{n=1}^{N_f} f(x_{i,n}) U_{i,n} + \sum_{n=1}^{N_f} \int_0^{f(x_{i,n})} f^{-1}(z) dz$$

$$\begin{aligned}
&= - \sum_{n=1}^{N_f} \sum_{m=1}^{N_f} \frac{1}{2} T_{nm} \tanh(x_{i,m}) \tanh(x_{i,n}) \\
&\quad - \sum_{n=1}^{N_f} \tanh(x_{i,n}) U_{i,n} + \sum_{n=1}^{N_f} \int_0^{\tanh(x_{i,n})} f^{-1}(z) dz,
\end{aligned}$$

where $U_{i,n} = \sum_{j \in \mathcal{N}} w_{ij} \times e^{(-\frac{(\theta_n - \theta_j^{\text{in}})^2}{2\sigma_{\text{in}}^2})}$. The derivation of the energy function can be deduced as follows:

$$\begin{aligned}
\dot{E}_i &= \sum_{n=1}^{N_f} \frac{\partial E_i}{\partial x_{i,n}} \dot{x}_{i,n} \\
&= \sum_{n=1}^{N_f} \left[- (1 - \tanh^2(x_{i,n})) \sum_{m=1}^{N_f} T_{nm} \tanh(x_{i,m}) \right. \\
&\quad \left. - U_{i,n} (1 - \tanh^2(x_{i,n})) + x_{i,n} (1 - \tanh^2(x_{i,n})) \right] \dot{x}_{i,n} \\
&= - \sum_{n=1}^{N_f} \left[- x_{i,n} + \sum_{m=1}^{N_f} T_{nm} \tanh(x_{i,m}) + U_{i,n} \right] \\
&\quad \times (1 - \tanh^2(x_{i,n})) \dot{x}_{i,n} \\
&= - \sum_{n=1}^{N_f} \dot{x}_{i,n}^2 (1 - \tanh^2(x_{i,n})) \leq 0,
\end{aligned}$$

where $\dot{x}_{i,n} = -x_{i,n} + \sum_{m=1}^{N_f} T_{nm} \tanh(x_{i,m}) + U_{i,n}$. According to LaSalle's invariance principle, this guarantees that for any given external input U_i , the neural activity of agent i will always converge to a stable equilibrium (an attractor state) where the energy E_i is at a local minimum. This ensures the reliability and robustness of the decision-making core in producing stable orientation outputs.

APPENDIX E: STABILITY ANALYSIS OF THE MULTIAGENT SYSTEM

1. Timescale separation and quasistationary approximation

We assume that the internal neurodynamics governed by Eq. (1) operate on a much faster timescale than the agents' dynamics. Therefore, at each instant of the slow-motion dynamics, the Hopfield network has effectively equilibrated to the attractor set anchored by the current sensory drive U_i .

In this quasisteady regime, while the recurrent connections sharpen the neural activity, the peak position (encoded heading) of the neural state x_i is effectively locked to the peak of the external input U_i . This approximation permits us to analyze the readout dynamics by focusing on the aggregate input profile:

$$x_i \approx U_i = \sum_{j \in \mathcal{N}_i} w_{ij} \times \eta_j.$$

2. Linear approximation of the neural readout

Under the quasistationary assumption (the Hopfield core equilibrates much faster than agent motion), the nonlinear softmax readout of agent i can be analyzed in a continuous feature limit. First, we write the input to the network as a continuous function of the angle α :

$$x_i(\alpha) = \sum_{j \in \mathcal{N}_i} w_{ij} g(\alpha - \theta_j),$$

where $g(u) = e^{-u^2/(2\sigma_{\text{in}}^2)}$. Invoking the soft-argmax method, we can obtain an approximation of θ_i^{out} as $\theta_i^{\text{out}} \approx \sum_{n=1}^{N_f} p_{i,n} \alpha_n$, where $p_{i,n} = \frac{e^{\beta x_{i,n}}}{\sum_{m=1}^{N_f} e^{\beta x_{i,m}}}$ ($\beta > 0$). The softmax readout gives

$$\theta_i^{\text{out}} = \frac{\int_{-\pi}^{\pi} \alpha e^{\beta x_i(\alpha)} d\alpha}{\int_{-\pi}^{\pi} e^{\beta x_i(\alpha)} d\alpha}.$$

In the strong-competition regime (β is sufficiently large), the integrals are dominated by the global maximizer α^* of $x_i(\alpha)$. According to Laplace's method, v_{max} can be approximated as

$$\theta_i^{\text{out}} = \alpha^* + O(1/\beta).$$

Thus, the problem reduces to finding the peak position α^* of the summed input signal. The peak position α^* is found by setting the first derivative of $x_i(\alpha)$ to zero:

$$\left. \frac{dx_i}{d\alpha} \right|_{\alpha=\alpha^*} = \sum_{j \in \mathcal{N}_i} w_{ij} g'(\alpha^* - \theta_j) = 0,$$

where $g'(u) = -(u/\sigma_{\text{in}}^2) e^{-u^2/(2\sigma_{\text{in}}^2)}$. Then, it can be deduced that

$$\sum_{j \in \mathcal{N}_i} w_{ij} (\alpha^* - \theta_j) e^{-(\alpha^* - \theta_j)^2/(2\sigma_{\text{in}}^2)} = 0.$$

If the neighbor headings $\{\theta_j\}$ lie within the kernel width ($|\alpha^* - \theta_j| \ll \sigma_{\text{in}}$ for all j), then $e^{-(\alpha^* - \theta_j)^2/(2\sigma_{\text{in}}^2)} \approx 1 + O((\alpha^* - \theta_j)^2/\sigma_{\text{in}}^2)$. Combining this with the Laplace estimate gives the leading-order approximation as

$$\sum_{j \in \mathcal{N}_i} w_{ij} (\alpha^* - \theta_j) \approx 0.$$

Since $\sum_j w_{ij} = 1$, we can deduce that $\alpha^* \approx \sum_j w_{ij} \theta_j$. Combining with the Laplace result yields

$$\theta_i^{\text{out}} = \sum_j w_{ij} \theta_j + O(1/\beta) + O((\Delta\theta/\sigma_{\text{in}})^2) + O(1/N_f),$$

where $\Delta\theta$ denotes the neighbor heading spread and $O(1/N_f)$ is the discretization error from replacing the finite feature set with the continuum. Combining the previous results, we can conclude that

$$\theta_i^{\text{out}} = \sum_{n=1}^{N_f} p_{i,n} \alpha_n \approx \sum_{j \in \mathcal{N}_i} w_{ij} \theta_j.$$

3. Stability of the dynamics

We next low-pass filter the readout to obtain a smooth agent heading ϕ_i . Let $\phi_i = \tau(-\phi_i + \theta_i^{\text{out}})$. Using the approximation

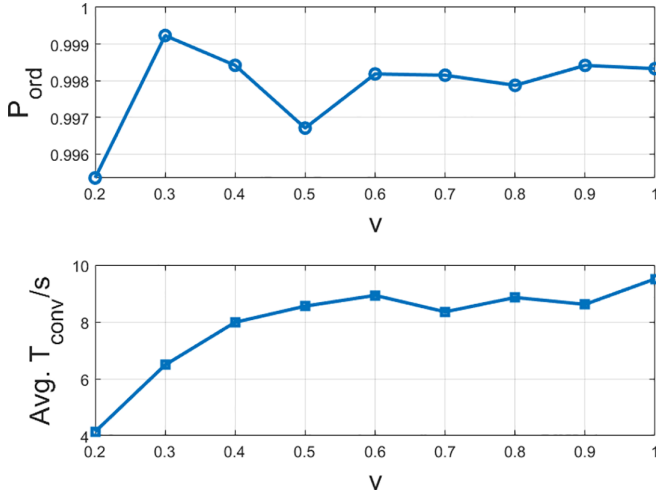


FIG. 11. Order parameter and convergence time of the swarm.

above and replacing θ_i by the continuous state ϕ_i yields the following closed form:

$$\begin{aligned}\dot{\phi}_i &= \tau(-\phi_i + \theta_i^{\text{out}}) \\ &\approx \tau\left(-\phi_i + \sum_{j \in \mathcal{N}_i} w_{ij} \times \theta_j\right) \\ &\approx \tau\left(-\phi_i + \sum_{j \in \mathcal{N}_i} w_{ij} \times \phi_j\right) \\ &= \tau \sum_{j \in \mathcal{N}_i} w_{ij}(-\phi_i + \phi_j),\end{aligned}$$

where, in the last step, we replace the instantaneous headings θ_j by the smooth states ϕ_j , which coincide in the quasistatic regime. Moreover, the weights satisfy $\sum_{j=1}^N w_{ij} = 1$. The compact form can be written as $\dot{\phi} = -\tau L\phi$, where $L = I - W$ is the Laplacian matrix and $W = [w_{ij}]$.

Although L is generally asymmetric, standard results in consensus control theory [43] guarantee that if the directed interaction graph contains a directed spanning tree, then L has a simple eigenvalue at zero with right eigenvector $\mathbf{1}$, and all other eigenvalues have positive real parts. As a result, the consensus subspace $\{\phi \in \mathbb{R}^N : \phi_1 = \dots = \phi_N\}$ is globally asymptotically stable for $\dot{\phi} = -\tau L\phi$, and the agent headings converge to a common value (a consensus heading) determined by the left eigenvector associated with the zero eigenvalue.

APPENDIX F: PARAMETER SWEEPS AND PERFORMANCE METRICS

1. Standard deviation of the final order parameter

We define the final order parameter in run k as the time-average of P_{ord} over the last 100 time steps and denote it by $P_{\text{ord}}^{(k)}$. The mean and standard deviation are $\bar{P}_{\text{ord}} = \frac{1}{M} \sum_{k=1}^M P_{\text{ord}}^{(k)}$ and $\sigma_{P_{\text{ord}}} = \sqrt{\frac{1}{M-1} \sum_{k=1}^M (P_{\text{ord}}^{(k)} - \bar{P}_{\text{ord}})^2}$.

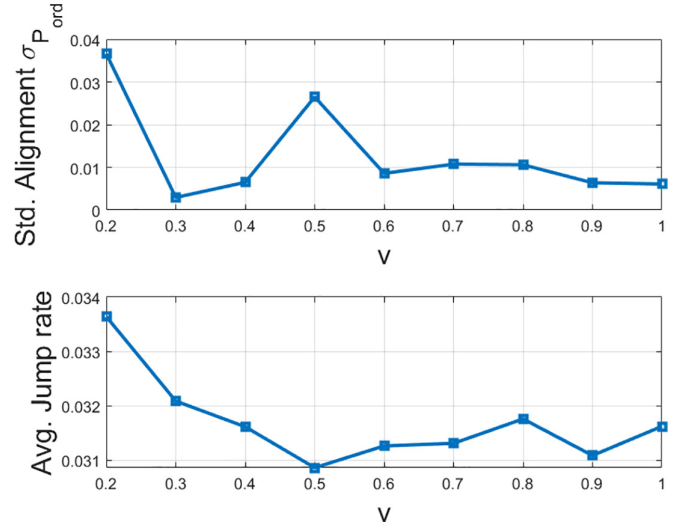


FIG. 12. Standard deviation of the final order parameter and rate of jump of the swarm.

2. Average jump rate of the swarm

A “jump” is counted when the minimal angular difference between an agent’s headings at two consecutive steps exceeds 30° . For simulation k , let $J^{(k)}$ denote the total number of such events over all agents and time steps. We define the jump rate as $\bar{J} = \frac{1}{M} \sum_{k=1}^M \frac{J^{(k)}}{N(T-1)}$.

3. Sharpness v

The exponent v modulates the sharpness of the excitation-inhibition profile in the ring-attractor network. To determine an appropriate range for the exponent v , we conducted extensive simulations by varying v over the interval $[0.2, 1.0]$. Each configuration was repeated 200 times under the following fixed parameters: $N = 50$, $v_0 = 2$, $L = 10$, and $R = 3$. The results are summarized in Figs. 11 and 12.

As shown in Fig. 11, the final order parameter P_{ord} remains consistently high across all values of v , with $P_{\text{ord}} > 0.99$, indicating reliable consensus formation. The average

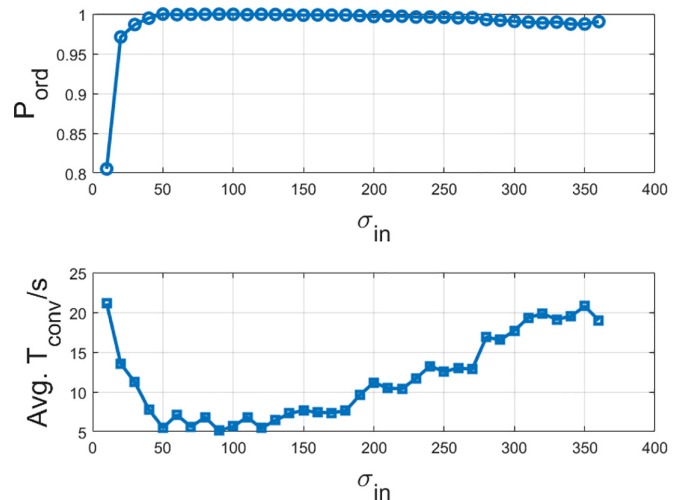


FIG. 13. Order parameter and convergence time of the swarm.

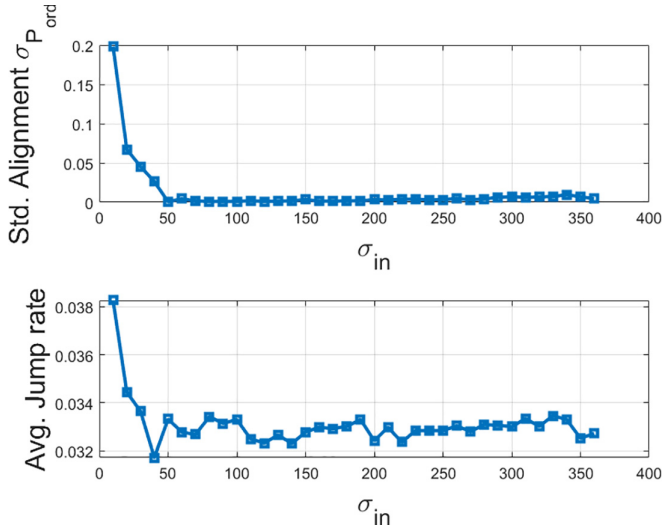


FIG. 14. Standard deviation of the final order parameter and rate of jump of the swarm.

convergence time T_{conv} exhibits a nonmonotonic trend: It increases with v in the range $v \in [0.2, 0.6]$, remains relatively stable for $v \in [0.6, 0.9]$, and slightly increases again at $v = 1.0$.

In Fig. 12, we observe that the standard deviation of the final order parameter remains low (generally below 0.02) for most values of v , with particularly stable alignment at $v = 0.1, 0.3$, and 1.0 . However, a notable increase in variability is seen at $v = 0.5$, where the standard deviation peaks at approximately 0.0505.

Regarding the average number of direction jumps, the system exhibits a clear downward trend as v increases from 0.2 to 0.5. This suggests that higher values of v lead to smoother directional transitions and less frequent abrupt changes in agent headings. However, beyond $v = 0.5$, the jump count stabilizes within a narrow range around 39, showing limited sensitivity to further increases in v .

Based on these observations, we conclude that the range $v \in [0.3, 0.7]$ achieves a desirable trade-off between convergence speed, directional smoothness, and network stability.

4. Input width σ_{in}

To evaluate the impact of the directional input width σ_{in} on the performance of the Hopfield-based swarm model, a series of simulations were conducted by varying σ_{in} from 10° to 360° , with each configuration repeated 50 times. The results are summarized in Figs. 13 and 14.

As shown in Fig. 13, the average final alignment P_{ord} increases rapidly from approximately 0.81 at $\sigma_{\text{in}} = 10^\circ$ to values exceeding 0.99 within the range $\sigma_{\text{in}} \in [50^\circ, 150^\circ]$, indicating a critical input bandwidth threshold for effective alignment. Beyond $\sigma_{\text{in}} > 200^\circ$, a slight decline in P_{ord} indicates reduced directional selectivity due to input dispersion.

The convergence time T_{conv} follows a U-shaped trend, reaching a minimum of 5 s around $\sigma_{\text{in}} = 60^\circ$, then increasing steadily for larger values. This highlights a trade-off: Narrow inputs lack robustness, while broad inputs slow convergence.

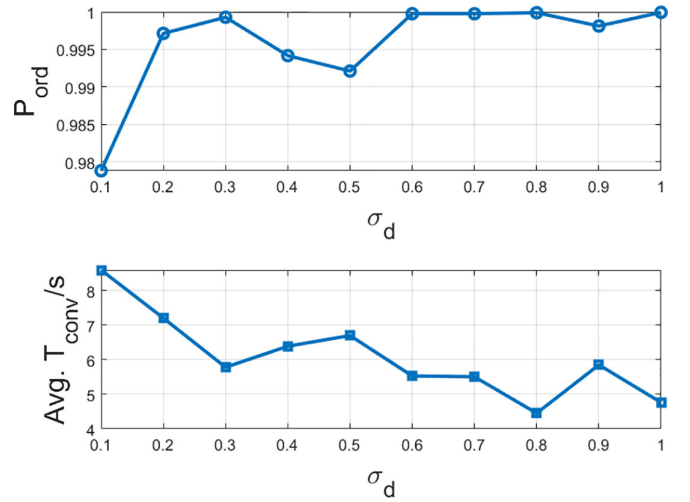


FIG. 15. Order parameter and convergence time of the swarm.

Figure 14 shows that the standard deviation of P_{ord} drops sharply near $\sigma_{\text{in}} = 40^\circ$ and remains below 0.005 for $\sigma_{\text{in}} \geq 50^\circ$, confirming alignment stability. The jump rate declines rapidly within $[10^\circ, 50^\circ]$ and stabilizes near 3.3%, reflecting smoother neural dynamics under moderate σ_{in} .

In summary, $\sigma_{\text{in}} \in [50^\circ, 150^\circ]$ yields the best balance between convergence speed and alignment reliability, underlining the importance of input encoding width in swarm decision-making based on continuous attractor networks.

5. Neighbor weighting scale σ_d

We investigate the effect of the neighbor weighting scale σ_d on swarm alignment using the same protocol. Consistent with the main text, Figs. 15 and 16 show that the range $\sigma_d \in [0.6L, 0.9L]$ achieves the best trade-off between convergence speed and alignment reliability.

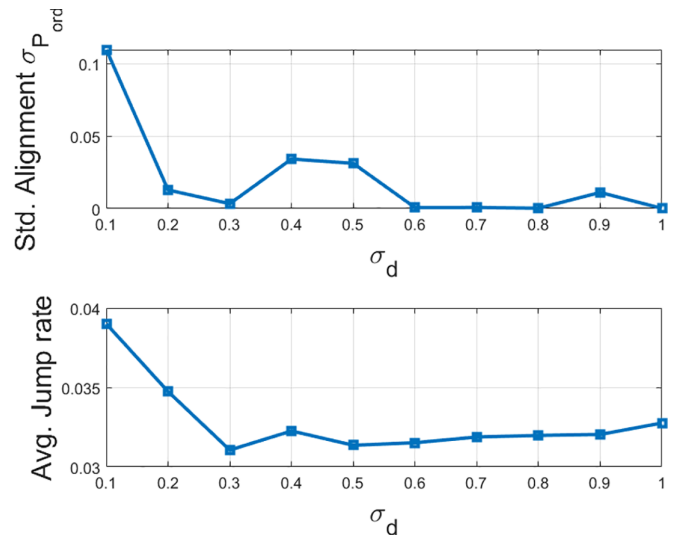


FIG. 16. Standard deviation of the final order parameter and rate of jump of the swarm.

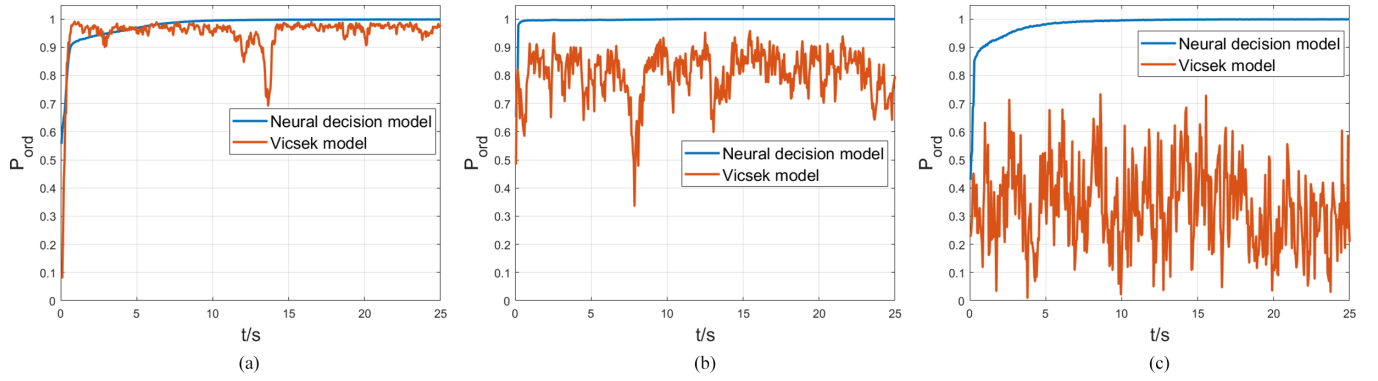


FIG. 17. Noise-robustness comparison with the Vicsek model. (a) At low noise ($a = 0.2$), both models attain high global order, but the neural decision model converges faster and displays smaller trial-to-trial variability. (b) At $a = 0.5$, the performance of the Vicsek model begins to degrade significantly, showing large fluctuations and a lower mean order. (c) At $a = 1.1$, the Vicsek model's coherence effectively collapses to near-disordered levels, whereas the neurodynamic model continues to sustain a high degree of global alignment.

APPENDIX G: ROBUSTNESS AGAINST STOCHASTIC NOISE: METHODS AND DETAILED RESULTS

1. Numerical protocol and baseline models

All noise-robustness experiments reported here use periodic toroidal boundary conditions. Unless stated otherwise, simulations employ the following parameters: $v_0 = 2$, interaction radius $R = 3$, time step $\Delta t = 0.01$, total simulation time $T_{\text{sim}} = 200$, and $M = 50$ independent realizations per parameter point. Initial positions are randomly sampled on the $L \times L$ domain and initial headings are drawn from $\theta_i(0) \in [0, 2\pi)$.

For fair comparison, we consider three baseline implementations: (1) a Vicsek model with alignment-only interactions, with parameters $N = 20$ and $L = 10$; (2) a Couzin model with full-interaction zones (repulsion/orientation/attraction), with parameters $r_r = 0.4$, $r_o = 2.6$, $r_a = 3.0$, $N = 20$, and $L = 15$; for comparison with this baseline, the neural model is augmented with Couzin-style repulsion and attraction applied at the velocity-command stage; and (3) a Glauber-type neurodynamic rule following Ref. [24], with parameters $v = 0.5$, $N = 20$, $L = 100$, $N_s = 100$, $v_0 = 2$, $\sigma = 2\pi/N_s$, $\beta = 10$, and $h_0 = 0.008$.

2. Discrete-time implementation

For numerical simulations, the agent kinematics are implemented in discrete time with step Δt as follows:

$$r_i(t + \Delta t) = r_i(t) + v_0 \Delta t \begin{bmatrix} \cos \theta_i^{\text{out}}(t) \\ \sin \theta_i^{\text{out}}(t) \end{bmatrix} + \delta_i(t), \quad (\text{G1})$$

where $\delta_i(t)$ is the additive per-step noise vector drawn independently at each time step from $\text{Unif}([-a, a]^2)$.

This discrete-time scheme puts the Vicsek, Couzin, and neural models on the same numerical footing and allows us to apply identical perstep noise statistics to all of them.

3. Comparative performance under different noise levels

Figure 5 reports ensemble time series for the neural decision model and three baselines (Vicsek, Couzin, and Glauber-based neurodynamic rule) under translational noise (uniform in $[-0.8, 0.8]^2$). Here, we further compare model responses across three representative noise amplitudes, $a = 0.2$,

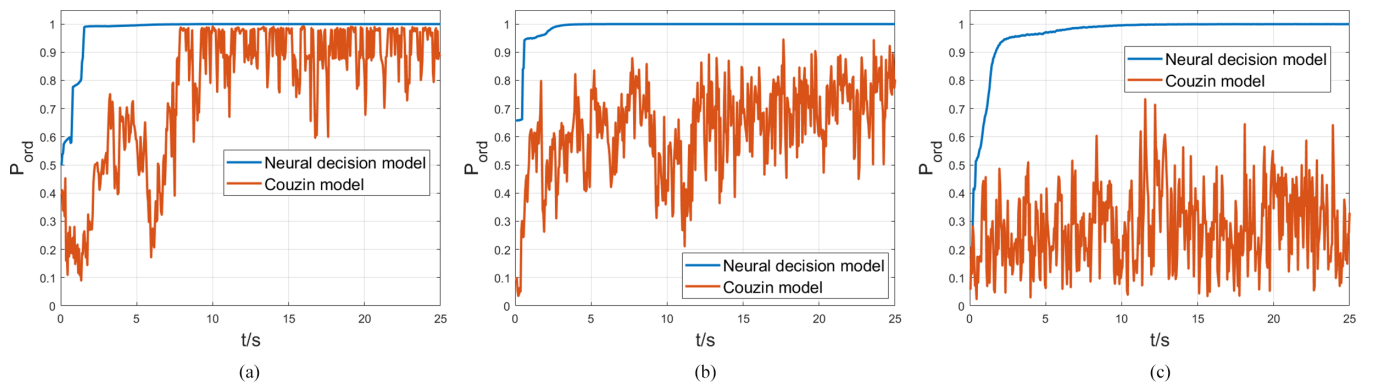


FIG. 18. Noise-robustness comparison with Couzin model. (a) At low noise ($a = 0.2$), both models converge to a highly ordered state, but the neurodynamic model reaches near-perfect alignment more rapidly and with smaller trial-to-trial variability. (b) At $a = 0.5$, the Couzin baseline begins to degrade: Its mean order decreases and temporal fluctuations increase. (c) At $a = 1.1$, the Couzin model exhibits large, persistent fluctuations and fails to maintain global coherence, whereas the neurodynamic swarm continues to sustain a high degree of alignment.

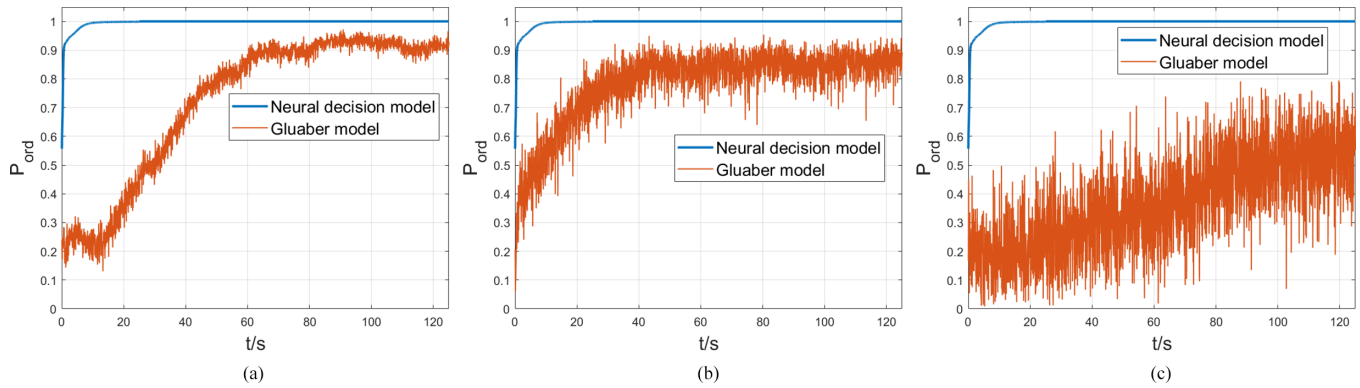


FIG. 19. Noise-robustness comparison with Glauber-based neurodynamic model. (a) Even at low noise ($a = 0.2$), the neural decision model converges to near-perfect alignment much faster and attains a more stable steady state. By contrast, the probabilistic Glauber model shows slower convergence and settles at a lower mean order with persistent trial-to-trial fluctuations. (b) At $a = 0.5$, the performance of the Glauber model further degrades, showing larger fluctuations. (c) At $a = 1.1$, the Glauber model is unable to establish robust global coherence, with its order parameter exhibiting large, persistent oscillations around a low mean. By comparison, the deterministic attractor readout preserves a high degree of alignment.

0.5, and 1.1. Figures 17–19 show the results. The neural decision model preserves near-perfect alignment across all tested

noise scales, whereas alignment-based rules lose coherence progressively as a increases.

- [1] D. J. T. Sumpter, *Collective Animal Behavior* (Princeton University Press, Princeton, NJ, 2010).
- [2] S. Camazine, J.-L. Deneubourg, G. Theraula, J. Sneyd, and N. R. Franks, *et al.*, *Self-Organization in Biological Systems* (Princeton University Press, Princeton, NJ, 2020).
- [3] T. S. Deisboeck and I. D. Couzin, Collective behavior in cancer cell populations, *BioEssays* **31**, 190 (2009).
- [4] L. Li, M. Nagy, G. Amichay, R. Wu, W. Wang, O. Deussen, D. Rus, and I. D. Couzin, Reverse engineering the control law for schooling in Zebrafish using virtual reality, *Sci. Rob.* **10**, eadq6784 (2025).
- [5] 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).
- [6] G. Palla, I. Derényi, I. Farkas, and T. Vicsek, Uncovering the overlapping community structure of complex networks in nature and society, *Nature (London)* **435**, 814 (2005).
- [7] T. Vicsek, A. Czirók, E. Ben-Jacob, I. Cohen, and O. Shochet, Novel type of phase transition in a system of self-driven particles, *Phys. Rev. Lett.* **75**, 1226 (1995).
- [8] 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).
- [9] P. Jentsch and C. F. Lee, New universality class describes Vicsek's flocking phase in physical dimensions, *Phys. Rev. Lett.* **133**, 128301 (2024).
- [10] T. Vicsek and A. Zafeiris, Collective motion, *Phys. Rep.* **517**, 71 (2012).
- [11] 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).
- [12] I. D. Couzin, Collective cognition in animal groups, *Trends Cognit. Sci.* **13**, 36 (2009).
- [13] C. Virág, G. Vásárhelyi, N. Tárcai, T. Szörényi, G. Somorjai, T. Nepusz, and T. Vicsek, Flocking algorithm for autonomous flying robots, *Bioinspir. Biomim.* **9**, 025012 (2014).
- [14] G. Vásárhelyi, C. Virág, G. Somorjai, T. Nepusz, A. E. Eiben, and T. Vicsek, Optimized flocking of autonomous drones in confined environments, *Sci. Robot.* **3**, eaat3536 (2018).
- [15] M. R. D'Orsogna, Y.-L. Chuang, A. L. Bertozzi, and L. S. Chayes, Self-propelled particles with soft-core interactions: Patterns, stability, and collapse, *Phys. Rev. Lett.* **96**, 104302 (2006).
- [16] M. Ballerini, N. Cabibbo, R. Candelier, A. Cavagna, E. Cisbani, I. Giardina, V. Lecomte, A. Orlandi, G. Parisi, A. Procaccini, *et al.*, Interaction ruling animal collective behavior depends on topological rather than metric distance: Evidence from a field study, *Proc. Natl. Acad. Sci. USA* **105**, 1232 (2008).
- [17] F. Ginelli and H. Chaté, Relevance of metric-free interactions in flocking phenomena, *Phys. Rev. Lett.* **105**, 168103 (2010).
- [18] D. L. Krongauz, A. Ayali, and G. A. Kaminka, Vision-based collective motion: A locust-inspired reductionist model, *PLoS Comput. Biol.* **20**, e1011796 (2024).
- [19] S. Bazazi, C. Buhl, J. J. Hale, M. L. Anstey, G. A. Sword, S. J. Simpson, and I. D. Couzin, Collective motion and cannibalism in locust migratory bands, *Curr. Biol.* **18**, 735 (2008).
- [20] S. Bazazi, F. Bartumeus, J. J. Hale, and I. D. Couzin, Intermittent motion in desert locusts: Behavioural complexity in simple environments, *PLoS Comput. Biol.* **8**, e1002498 (2012).
- [21] C. 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).
- [22] H. Chang, S. Cassau, J. Krieger, X. Guo, M. Knaden, L. Kang, and B. S. Hansson, A chemical defense deters cannibalism in migratory locusts, *Science* **380**, 537 (2023).

- [23] Z. Gao and I. D. Couzin, Swarming transitions of self-propelled particles with anisotropic social interactions, *Phys. Rev. Res.* **7**, 023144 (2025).
- [24] S. Sayin, E. Couzin-Fuchs, I. Petelski, Y. Günzel, M. Salahshour, C.-Y. Lee, J. M. Graving, L. Li, O. Deussen, G. A. Sword, *et al.*, The behavioral mechanisms governing collective motion in swarming locusts, *Science* **387**, 995 (2025).
- [25] C. A. Yates, R. Erban, C. Escudero, I. D. Couzin, C. 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).
- [26] F. Ginelli, The physics of the Vicsek model, *Eur. Phys. J.: Spec. Top.* **225**, 2099 (2016).
- [27] V. Holubec, D. Geiss, S. A. M. Loos, K. Kroy, and F. Cichos, Finite-size scaling at the edge of disorder in a time-delay Vicsek model, *Phys. Rev. Lett.* **127**, 258001 (2021).
- [28] I. D. Couzin and J. Krause, Self-organization and collective behavior in vertebrates, in *Advances in the Study of Behavior*, edited by P. J. B. Slater, J. S. Rosenblatt, C. T. Snowden, and T. J. Roper (Academic Press, San Diego, CA, 2003), Vol. 32, pp. 1–75.
- [29] Y. Katz, K. Tunstrøm, 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).
- [30] K.-K. Oh, M.-C. Park, and H.-S. Ahn, A survey of multi-agent formation control, *Automatica* **53**, 424 (2015).
- [31] W. Ren, R. W. Beard, and E. M. Atkins, Information consensus in multivehicle cooperative control, *IEEE Control Syst. Mag.* **27**, 71 (2007).
- [32] T. Xue, X. Li, P. Grassberger, and L. Chen, Swarming transitions in hierarchical societies, *Phys. Rev. Res.* **2**, 042017(R) (2020).
- [33] D. Levis, I. Pagonabarraga, and B. Liebchen, Activity induced synchronization: Mutual flocking and chiral self-sorting, *Phys. Rev. Res.* **1**, 023026 (2019).
- [34] J. J. Hopfield, Neural networks and physical systems with emergent collective computational abilities, *Proc. Natl. Acad. Sci. USA* **79**, 2554 (1982).
- [35] C. Eliasmith, A unified approach to building and controlling spiking attractor networks, *Neural Comput.* **17**, 1276 (2005).
- [36] M. Khona and I. R. Fiete, Attractor and integrator networks in the brain, *Nat. Rev. Neurosci.* **23**, 744 (2022).
- [37] S. S. Kim, H. Rouault, S. Druckmann, and V. Jayaraman, Ring attractor dynamics in the *Drosophila* central brain, *Science* **356**, 849 (2017).
- [38] M. Penrose, *Random Geometric Graphs* (Oxford University Press, Oxford, 2003), Vol. 5.
- [39] J. Dall and M. Christensen, Random geometric graphs, *Phys. Rev. E* **66**, 016121 (2002).
- [40] L. Lamport, R. Shostak, and M. Pease, The Byzantine Generals Problem, in *Concurrency: The Works of Leslie Lamport*, edited by D. Malkhi (Association for Computing Machinery, New York, NY, 2019), pp. 203–226.
- [41] See Supplemental Material at <http://link.aps.org/supplemental/10.1103/xwdd-hjhn> for movie.mp4.
- [42] R. J. Glauber, Time-dependent statistics of the Ising model, *J. Math. Phys.* **4**, 294 (1963).
- [43] W. Ren and R. W. Beard, Consensus seeking in multiagent systems under dynamically changing interaction topologies, *IEEE Trans. Autom. Control* **50**, 655 (2005).