Collective motion from optic flow and retinal cues: an omniscient baseline and a visual perception pipeline

Luka Brodnik[a]

[a]University of Ljubljana, Faculty of Computer and Information Science

Self-propelled particle models reproduce swarming, schooling and milling, but they assume that every agent knows the exact positions and headings of its neighbours. Castro, Ruffier and Eloy (Phys. Rev. Research 6, 023016, 2024) replaced this knowledge with what an eye can measure: the angular width of the shades in the visual field and their optic flow. We re-implement their omniscient and visual models in a tested Python package. The visual pipeline builds occluding shades, estimates distance from shade width, tracks shades between time steps, removes the agent's own rotation and converts optic flow and its divergence into relative velocities. The cues are exact for isolated neighbours. In a group of 50, however, 58% of the shades are merged silhouettes of several agents, which look on average 24% closer than their nearest member. At the published gains, the omniscient model reproduces all three phases, and its phase diagrams show the expected role of the blind angle: with a rear blind spot, milling fills the region between swarming and schooling; without it, milling almost disappears. The visual model swarms, schools and switches spontaneously from milling to schooling. At the published milling gains, however, it mills only intermittently, because its milling region is shifted towards weaker attraction. We conclude with the plan for the full reproduction and for two extensions, unequal body sizes and a finite visual range.

collective motion | optic flow | visual perception | self-propelled particles | milling

Introduction

Animal groups move in a few recurring patterns: cohesive but disordered swarms, polarised schools and rotating mills (1, 2). Minimal self-propelled particle models explain them with local rules. Agents align with neighbours inside a radius (3), add repulsion and attraction zones (4, 5), or follow rules fitted to tracked fish (6). All of these models assume that an agent knows the relative positions and headings of its neighbours exactly. An animal or a robot has no such knowledge. It sees neighbours only as dark shapes on its retina, nearby individuals hide those behind them, and distance and heading must be inferred from the image.

Vision has been brought into the models in several ways. Interaction networks built from what fish can actually see predict information transfer better than metric or topological ones (7). Steering by the projection of the flock on the retina produces flocks of marginal opacity (8), and collective motion can emerge from a purely binary visual field (9). Castro, Ruffier and Eloy (10) kept the structure of a classical attraction–alignment model but computed both terms from retinal cues only. The angular width of a shade gives an apparent distance. Its optic flow, the motion of the image on the retina, gives the relative velocity; the divergence of the flow is the looming cue that animals use to time collisions (11). This visual model reproduces swarming, schooling and milling and shows noise-driven switching between milling and schooling (10).

The project has three aims: to reproduce both models and their phase diagrams, to break the visual model's assumption that all agents have the same size (so that shade width encodes distance), and to limit how far agents can see. This first report lays the foundation: a simulation framework, the omniscient model and its three phases, and a visual perception pipeline that we test against ground truth.

Methods

Model. N discs of radius a move at constant speed U with heading θ_i and direction $\mathbf{e}_i = (\cos\theta_i, \sin\theta_i)$:

$$\dot{\mathbf{x}}_i = U\mathbf{e}_i, \qquad \dot{\theta}_i = k_\odot\, \omega_{\odot,i} + k_\parallel\, \omega_{\parallel,i} + k_\eta\, \eta_i(t), \tag{1}$$

with white rotational noise η_i and $a = U = 1$. Both turning terms are normalised averages, $\langle \sum_j f_j g_j \rangle \equiv \sum_j f_j g_j / \sum_j |f_j|$, so $|\omega| \le 1$. The weight $b_\epsilon(\varphi) = 1 + \epsilon\cos\varphi$, where φ is the retinal angle measured from the heading, models a rear blind angle: vision is isotropic for $\epsilon = 0$, and the agent is blind straight behind for $\epsilon = 1$. There is no explicit collision avoidance (10). In the *omniscient* model, with distance r_{ij} and bearing φ_{ij},

$$\omega_\odot^{\mathrm{omni}} = \Big\langle \sum_j r_{ij}\, b_\epsilon(\varphi_{ij}) \sin\varphi_{ij} \Big\rangle, \qquad \omega_\parallel^{\mathrm{omni}} = \Big\langle \sum_j r_{ij}^{-2}\, b_\epsilon(\varphi_{ij})\, \sin(\theta_j - \theta_i) \Big\rangle. \tag{2}$$

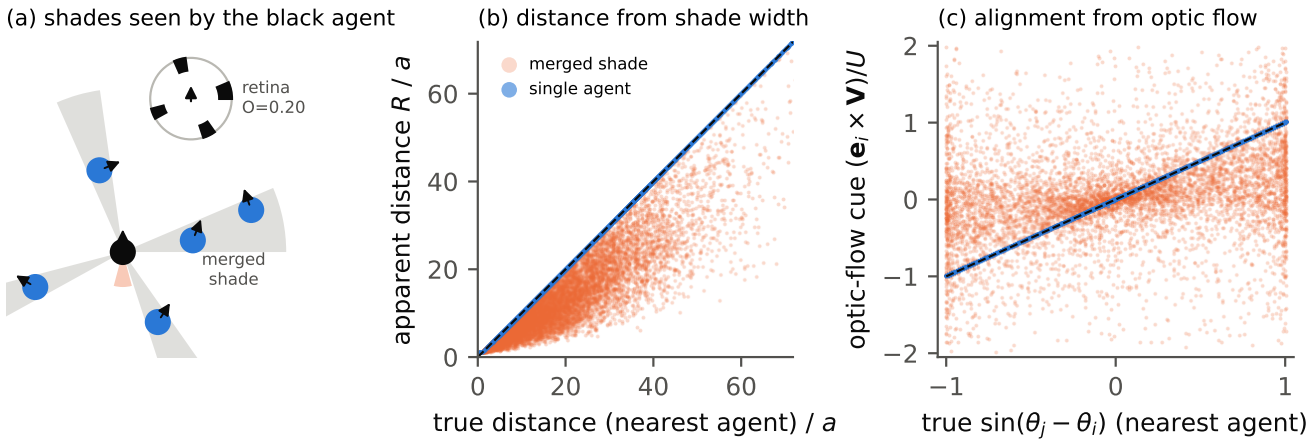


Fig. 1. Visual perception. (a) The black agent (heading up) sees five neighbours as four shades: two neighbours merge into one shade. Inset: its retina. The orange sector marks the blind direction for $\epsilon = 1$. (b) Apparent distance $\mathcal{R} = a/\sin(\Delta\varphi/2)$ and (c) the optic-flow alignment cue, both against ground truth for the nearest agent in each shade, from 45 states of the visual-model runs. Blue: single-agent shades; orange: merged shades; dashed: identity.

Visual field. A disc at distance r covers the retinal interval $[\varphi - h, \varphi + h]$ with $h = \arcsin(a/r)$. The union of these intervals is the binary visual field $V_i(\varphi)$. Overlapping intervals merge into one *shade* k with centre φ_k and width $\Delta\varphi_k$, so occlusion is built in. We compute the union by sorting the intervals and taking a running maximum of their ends, then join the first and last shade across $\pm\pi$. A shade read as a single disc has the apparent distance $\mathcal{R}_k = a/\sin(\Delta\varphi_k/2)$. This is exact for an isolated neighbour and too small for a merged shade.

Optic flow. Every step, an agent rotates its previous shades by its own heading change $\delta\theta_i$ (derotation) and pairs each current shade with a previous one, minimising the apparent displacement $|\mathcal{R}_k\mathbf{e}(\varphi_k) - \mathcal{R}_l^-\mathbf{e}(\varphi_l^- - \delta\theta_i)|$ with the Hungarian algorithm (12). Pairs that move farther than the largest possible relative displacement $2U\delta t$ are dropped; these are shades that merge, split, appear or disappear. For a tracked shade with derotated shift Δ_k, the optic flow and its divergence are

$$O_k = \frac{\mathcal{R}_k^- \sin\Delta_k}{\mathcal{R}_k\,\delta t}, \qquad D_k = -\frac{\mathcal{R}_k - \mathcal{R}_k^-\cos\Delta_k}{\mathcal{R}_k\,\delta t}, \tag{3}$$

which tend to $\dot\varphi_k$ and to the looming rate $-\dot{\mathcal{R}}_k/\mathcal{R}_k$ as $\delta t \to 0$. The relative velocity has components $V_r = -\mathcal{R}D$ and $V_\varphi = \mathcal{R}O$ (10). With Eq. 3 these components reproduce the displacement of an isolated neighbour over one step exactly.

Visual model. The turning rates are retinal integrals (10),

$$\omega_\odot^{\text{visu}} = \Big\langle \int_{-\pi}^{\pi} \mathcal{R}^2 b_\epsilon \sin\varphi\,\mathrm{d}\varphi \Big\rangle, \qquad \omega_\parallel^{\text{visu}} = \Big\langle \int_{-\pi}^{\pi} \frac{\mathbf{e}_i \times \mathbf{V}}{U\mathcal{R}}\, b_\epsilon\,\mathrm{d}\varphi \Big\rangle. \tag{4}$$

Within a shade, $\mathcal{R}$ and $\mathbf{V}$ are constant and $\mathbf{e}_i \times \mathbf{V} = V_r\sin\varphi_k + V_\varphi\cos\varphi_k$, so we evaluate the integrals exactly. With $h_k = \Delta\varphi_k/2$, $\int b_\epsilon\,\mathrm{d}\varphi = 2(h_k + \epsilon\cos\varphi_k\sin h_k)$ and $\int b_\epsilon\sin\varphi\,\mathrm{d}\varphi = 2\sin h_k\sin\varphi_k(1+\epsilon\cos\varphi_k\cos h_k)$. For a distant isolated disc, $\mathcal{R}^2\sin h_k = a\mathcal{R}$, and Eq. 4 reduces to Eq. 2.

Simulations. We integrate Eq. 1 with the Euler–Maruyama scheme (13), $\theta^{n+1} = \theta^n + \delta t(k_\odot\omega_\odot + k_\parallel\omega_\parallel) + k_\eta\sqrt{\delta t}\,\xi^n$ with $\xi^n \sim \mathcal{N}(0,1)$ and $\mathbf{x}^{n+1} = \mathbf{x}^n + U\delta t\,\mathbf{e}(\theta^{n+1})$. We use $N = 50$, $\delta t = 0.1$, $k_\eta = 0.01$ and $T = 5000$, starting from random, non-overlapping positions in a square of side aN with random headings (10). We measure the polarisation $P = |\langle\mathbf{e}_i\rangle|$; the milling parameter $M = |\langle\mathbf{y}_i \times \mathbf{e}_i\rangle|$, where $\mathbf{y}_i$ is the unit vector from the group centre to agent i; the opacity $O = \langle\frac{1}{2\pi}\int V_i\,\mathrm{d}\varphi\rangle$; and the RMS group radius S. All are averaged over the last 1000 time units. A run is schooling if $P \geq 0.5$, milling if $M \geq 0.5$, mixed if both hold and swarming otherwise.

Results

The visual cues are exact for isolated neighbours. Tests compare the visual field with a brute-force rasterisation of the retina, including shades that wrap around $\pm\pi$. For two agents, a parallel neighbour gives $O = D = 0$ and a head-on neighbour gives $V_r = -2U$ and $D > 0$. For every bearing and heading difference, $(\mathbf{e}_i \times \mathbf{V})/U = \sin(\theta_j - \theta_i)$ to machine precision, also while the observer turns. Fig. 1 applies the pipeline to 45 states of the visual-model runs (22 652 shades). Only 42% of the shades belong to a single agent. For these shades, $\mathcal{R}$ equals the true distance and the alignment cue lies on the identity line (error < 0.05). The remaining 58% are merged silhouettes. They look on average 24% closer than their nearest member (median $\mathcal{R}/r = 0.76$), and their alignment cue correlates only weakly with the true heading difference ($\rho = 0.40$). Of all shades, 86% are tracked from one step to the next.

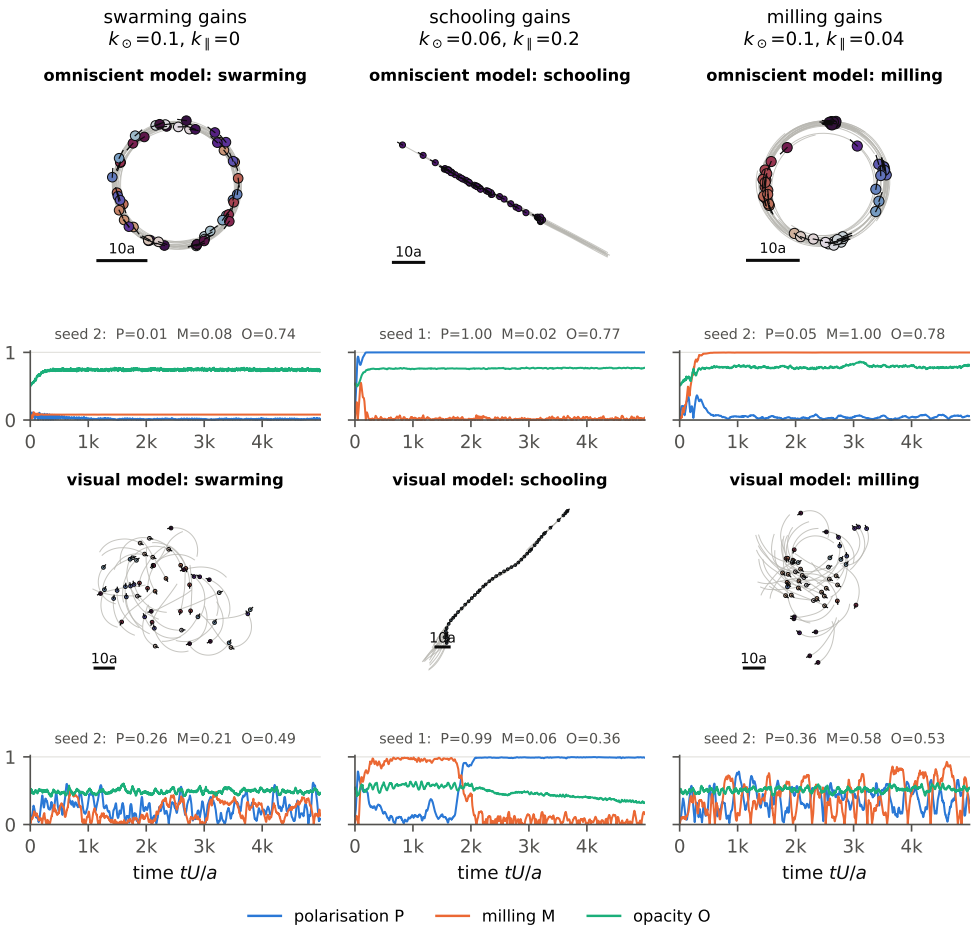


Fig. 2. Both models at the swarming, schooling and milling gains of (10) ($\epsilon = 1$). Each panel shows the run that reached the majority phase of its three seeds and is closest to their mean (seed given above the time series). Each column shows a gain pair: the final configuration (discs coloured by heading, trails over the last 25 time units, scale bar $10a$) above the order parameters over time; the omniscient model is on top and the visual model below.

Three phases. Fig. 2 and Table 1 show both models at the gains of Fig. 2 in (10) ($\epsilon = 1$, three seeds). The omniscient model reaches the intended phase in every run. Without alignment, agents circle the group centre in both directions on a ring of radius $\approx U/k_\odot = 10a$, so P and M stay low. With strong alignment they form an elongated school with $P = 1.00$, and at $k_\odot = 0.1$, $k_\parallel = 0.04$ a mill with $M = 1.00$. The visual model swarms and schools at the same gains. Its groups are about twice as wide and less opaque ($O = 0.3$–0.5 against 0.74–0.79). In one schooling run, the visual group first mills and then switches to schooling at $t \approx 1800$. This is the noise-induced transition reported in (10). At the milling gains, the visual model mills only intermittently: M fluctuates between 0 and 0.8, averaging 0.45 ± 0.15 (two seeds milling, one swarming).

Table 1. Steady-state order parameters (mean $\pm$ s.d. over three seeds; $\epsilon = 1$, $N = 50$, last 1000 of $T = 5000$ time units) and CPU time per run.

$(k_\odot, k_\parallel)$	model	P	M	O	S/a	classified as	time
(0.10, 0)	omniscient	0.01 ± 0.00	0.09 ± 0.05	0.74	12.7	3 swarming	11 s
	visual	0.28 ± 0.07	0.18 ± 0.04	0.49	24.3	3 swarming	125 s
(0.06, 0.2)	omniscient	1.00 ± 0.00	0.02 ± 0.00	0.79	12.9	3 schooling	10 s
	visual	1.00 ± 0.00	0.07 ± 0.01	0.31	43.2	3 schooling	110 s
(0.10, 0.04)	omniscient	0.05 ± 0.01	1.00 ± 0.00	0.79	11.8	3 milling	11 s
	visual	0.40 ± 0.05	0.45 ± 0.15	0.52	22.1	2 milling, 1 swarming	119 s

Phase diagrams. Fig. 3 maps the phases over $k_\odot, k_\parallel \in [0, 0.2]$. In the omniscient model with $\epsilon = 0$, schooling covers almost the whole diagram (104 of 121 cells). Swarming is confined to $k_\parallel = 0$ and very weak attraction, and milling appears in only three cells, at strong attraction and weak alignment. With $\epsilon = 1$, milling becomes the dominant phase (60 cells). It fills the region between swarming ($k_\parallel \lesssim 0.02$) and schooling, which now requires roughly $k_\odot \lesssim 0.6\, k_\parallel$. The coarse visual maps show the

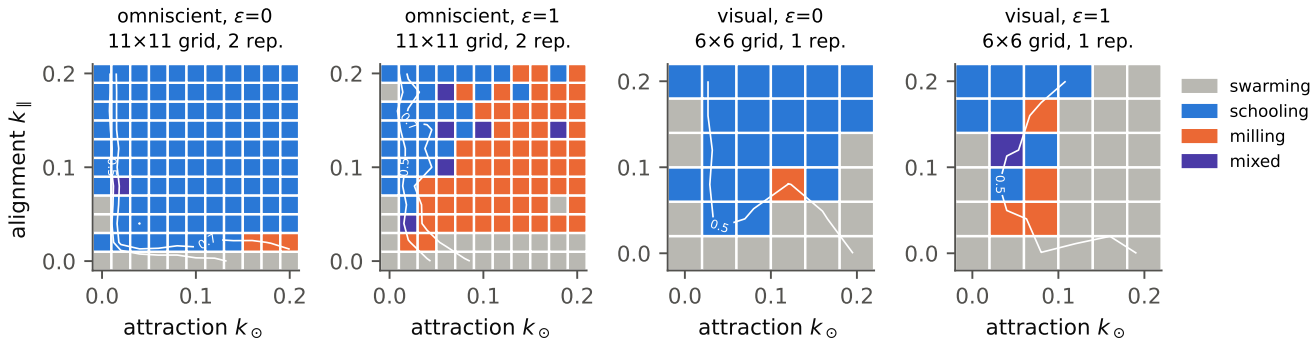


Fig. 3. Phases over the attraction and alignment gains ($N = 50$, $k_\eta = 0.01$, $T = 5000$), classified from P and M averaged over the last 1000 time units and over the repeats. White contours: opacity O. The visual-model maps are a coarse preview with one run per cell.

same organisation. For $\epsilon = 0$, schooling dominates (21 of 36 cells, one milling cell). For $\epsilon = 1$, the visual model mills strongly (M up to 0.94) at weak attraction, $k_\odot = 0.04$–0.08, and moderate alignment. Unlike the omniscient groups, however, the visual groups swarm at strong attraction ($k_\odot \geq 0.12$). The opacity of the visual groups never exceeds 0.67, against 0.93–0.98 in the omniscient model.

Discussion

The omniscient model behaves as reported in (10): all three phases appear at the published gains, and the rear blind angle creates milling. This validates the framework and gives a baseline for the extensions. The visual pipeline is exact whenever a shade belongs to a single agent, so its errors come only from occlusion. Those errors are large: in a group of 50, most shades are merged silhouettes that look too close and carry an unreliable alignment cue. The visual model still swarms, schools and switches from milling to schooling, so the collective behaviour tolerates these errors. The errors probably also make the visual groups looser and less opaque than the omniscient ones. The coarse visual phase map also explains the weak milling at the published milling gains: the visual milling region lies at weaker attraction than the omniscient one, and $k_\odot = 0.1$ is on its border with swarming. Our implementation differs from the reference code in three details: the shade integrals are exact, shades are matched by apparent displacement, and an overlapping agent covers half of the visual field. Each of these could shift that border. The low opacity of the visual groups agrees with the ceiling of about 0.7 reported in (10).

Plan. Report 2 will complete the reproduction: full-resolution phase diagrams of both models, with a direct comparison against the reference code at the milling border, for $\epsilon = 0$ and 1, and switching statistics between milling and schooling as a function of noise. The perception code already accepts a body radius per agent. We will add a visual range (a distance cutoff, contrast that fades with distance, or a minimum resolvable angle) to the shade computation. Report 3 will study (i) groups with a spread of body sizes, in which large agents look closer and small agents farther than they are, including whether agents sort by size; (ii) the effect of a finite visual range on milling, schooling and group fragmentation; and (iii) their combination. If large agents are seen from farther away than small ones, the visibility network becomes directed, and large agents may lead.

Data, Materials, and Software Availability. Code, data and figures are available in the project repository, https://github.com/lBrodnikl/SV2026_groupL, whose file `INSTRUCTIONS.md` explains how to regenerate every result. The reference implementation of (10) is at https://gitlab.com/FlokingByEyeDiego/flokingbyeye.

ACKNOWLEDGMENTS. We thank D. Castro, F. Ruffier and C. Eloy for making their code public.

1. T Vicsek, A Zafeiris, Collective motion. *Phys. Reports* **517**, 71–140 (2012).
2. K Tunstrøm, et al., Collective states, multistability and transitional behavior in schooling fish. *PLoS Comput. Biol.* **9**, e1002915 (2013).
3. T Vicsek, A Czirók, E Ben-Jacob, I Cohen, O Shochet, Novel type of phase transition in a system of self-driven particles. *Phys. Rev. Lett.* **75**, 1226–1229 (1995).
4. CW Reynolds, Flocks, herds and schools: A distributed behavioral model. *ACM SIGGRAPH Comput. Graph.* **21**, 25–34 (1987).
5. ID Couzin, J Krause, R James, GD Ruxton, NR Franks, Collective memory and spatial sorting in animal groups. *J. Theor. Biol.* **218**, 1–11 (2002).
6. DS Calovi, et al., Swarming, schooling, milling: phase diagram of a data-driven fish school model. *New J. Phys.* **16**, 015026 (2014).
7. A Strandburg-Peshkin, et al., Visual sensory networks and effective information transfer in animal groups. *Curr. Biol.* **23**, R709–R711 (2013).
8. DJG Pearce, AM Miller, G Rowlands, MS Turner, Role of projection in the control of bird flocks. *Proc. Natl. Acad. Sci.* **111**, 10422–10426 (2014).
9. R Bastien, P Romanczuk, A model of collective behavior based purely on vision. *Sci. Adv.* **6**, eaay0792 (2020).
10. D Castro, F Ruffier, C Eloy, Modeling collective behaviors from optic flow and retinal cues. *Phys. Rev. Res.* **6**, 023016 (2024).
11. DN Lee, A theory of visual control of braking based on information about time-to-collision. *Perception* **5**, 437–459 (1976).
12. HW Kuhn, The Hungarian method for the assignment problem. *Nav. Res. Logist. Q.* **2**, 83–97 (1955).
13. PE Kloeden, E Platen, *Numerical Solution of Stochastic Differential Equations.* (Springer, Berlin), (1992).