amU
Aix Marseille Université

NNT : 2025AIXM0016

THÈSE DE DOCTORAT

Soutenue à AMU — Aix-Marseille Université
le 13 Janvier 2024 par

Diego CASTRO

Flocking by eye:
Bioinspired flocking using only visual cues.

Discipline
Sciences du Mouvement Humain

École doctorale
ED 463
SCIENCES DU MOUVEMENT HUMAIN

Laboratoire/Partenaires de recherche
Institut des Sciences du Mouvement Etienne Jules Marey
Institut de Recherche sur les Phénomènes Hors Équilibre – I.R.P.H.E

Composition du jury

Prof. Eva KANSO Profesor & Zohrab A. Kaprielian Fellow in Engineering, University of Southern California, Los Angeles	Rapporteure
Dr. Ramiro GODOY DIANA Directeur de recherche CNRS, PMMH, ESPCI Paris	Rapporteur
Dr. Lara BRIÑÓN ARRANZ Maîtresse de conférence, Gipsa-Lab, Institut Polytechnique de Grenoble	Examinatrice
Dr. Aurore LOISY Chargée de Recherche CNRS, IRPHÉ, Aix-Marseille Université	Examinatrice
Dr. Guy THERAULAZ Directeur de recherche CNRS, CRCA,Université Toulouse 3 Paul Sabatier	Président du jury
Dr. Franck RUFFIER Directeur de Recherche CNRS, ISM, CNRS/Aix-Marseille Université	Directeur de thèse
Prof. Christophe ELOY Professeur à l'École Centrale Marseille, IRPHÉ, Aix-Marseille Université	Co-directeur de thèse

INSTITUT DES SCIENCES DU MOUVEMENT ETIENNE JULES MAREY

irphé Institut de Recherche sur les Phénomènes Hors Equilibre

cnrs

Affidavit

I, undersigned, Diego Mauricio Castro Rondon, hereby declare that the work presented in this manuscript is my own work, carried out under the scientific supervision of Franck Ruffier and Christophe Eloy, in accordance with the principles of honesty, integrity and responsibility inherent to the research mission. The research work and the writing of this manuscript have been carried out in compliance with both the french national charter for Research Integrity and AMU charter on the fight against plagiarism.

This work has not been submitted previously either in this country or in another country in the same or in a similar version to any other examination body.

Marseille, March 21, 2025



This work is licensed under Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International Public License

List of Publications

Journal Papers

1. Modeling collective behaviors from optic flow and retinal cues. Castro, D., Ruffier, F., & Eloy, C. (2024). "Modeling collective behaviors from optic flow and retinal cues." *Physical Review Research*, 6(2), 023016. DOI: 10.1103/PhysRevResearch.6.023016

2. Visual collective behaviors on spherical robots. Castro D., Eloy, C., & Ruffier, F. (2025). "Visual collective behaviors on spherical robots." *Biomimetics & Bioinspircation*, 20(2), 026006. DOI: 10.1088/1748-3190/adaab9

Posters

1. Journée des Jeunes Chercheurs en Robotique (JJCR) 2023, Moliets (Landes, France).

Other Communication

1. Journée de l'École Doctorale 2023, (Montpelier, France), 2éme place presentations orales.

Résumé et mots clés

Les mouvements collectifs sont omniprésents parmi de nombreuses espèces d'animaux, et ce quelle que soit leur taille ou mode de locomotion. Cependant, les modèles existants reposent sur les particules autopropulsées, dans lesquelles chaque particule possède des mesures artificielles "omniscientes" des distances et des vitesses de ses voisins. La disponibilité de ces mesures artificialise la perception des animaux de façon irréaliste, comme si un animal pouvait connaître les positions exactes de tous ses congénères membres d'un même groupe. En réalité, les animaux utilisent la vision comme base de leur comportement collectif.

L'objectif de cette thèse est de développer un modèle de mouvement collectif basé sur la perception visuelle des animaux. Dans cette étude, chaque individu voit les autres individus sous forme de sphères. Ainsi, chaque individu n'est pas identifié, mais il est perçu seulement pour leur projection sur la rétine panoramique de chaque individu. Les informations visuelles utilisées par les individus sont une combinaison de la taille optique d'un ou d'un amas d'individus, sa position rétinienne, son mouvement apparent (flux optique de translation), son expansion/contraction optique (divergence du flux optique). De ces informations visuelles naissent deux règles qui sont mises en œuvre et appliquées à tous les individus, à savoir l'attraction et l'alignement.

La première règle, l'attraction, basée la taille optique des individus telle que perçue par chaque individu dans son champ visuel : cette règle reflète la tendance des individus à rester ensemble. Sans attraction, les individus s'éloigneraient naturellement les uns des autres.

La seconde règle est l'alignement, basée sur le flux optique. Une distinction est faite entre le flux optique de translation, qui correspond au défilement optique des objets rétiniens, et la divergence du flux optique, qui correspond à la variation de taille optique des objets rétiniens. L'alignement reflète la tendance du groupe à se déplacer dans la même direction. Sans alignement, les individus du groupe auraient du mal à se suivre.

Ce modèle visuel comble l'écart entre les approches traditionnelles basées sur la physique des particules et la perception visuelle et il a permis de reproduire les principaux comportements collectifs :

- le bourdonnement (swarming), comme celui d'un groupe d'abeilles ou de moustiques qui volent ensemble dans une même zone,
- un banc ou une volée (schooling), comme un groupe d'oiseaux qui volent ensemble d'un endroit à l'autre, et,

- le vortex (milling), comme une multitude de poissons qui tournent en rond dans un mouvement tourbillonnant.

De plus, ce modèle visuel a été complété avec succès par l'introduction des contraintes nécessaires pour un essaim robotisé — un délai, une ancre et une règle d'évitement —. Pour la première fois, le comportement mise en œuvre sur des robots (ici sphériques) sont très semblables au comportement de particules autopropulsées en simulation suivant ce même modèle visuel complété. Ces similitudes ont été observées non seulement dans les comportements reproduits, mais aussi dans leur distribution dans l'espace des paramètres d'alignement et d'attraction. Ces robots sphériques ont un comportement est assez proche de celui des particules autopropulsées traditionnelles utilisées pour simuler le comportement collectif.

En résumé, ma thèse de doctorat propose un cadre biologiquement plausible pour comprendre le mouvement collectif avec un parallèle constant entre les méthodes traditionnelles basées sur les particules autopropulsées et leur mise en œuvre dans des robots ayant une cinématique similaire.

Mots clés: Mouvement Collectif, Information visuelle, Bioinspiré, Robotique, Vision, Flux Optique

Abstract and keywords

Collective movements are ubiquitous among many animal species, whatever their size or mode of locomotion. Traditionally, collective behaviors models rely on self-propelled particles, where each had "omniscient" knowledge of its neighbor's distances and velocities. The availability of theses artificial measurements, oversimplifies animal perception in an unrealistic way, as if an animal could know the exact positions of all individual of the same group. Animals would rather use their vision during their collective behavior.

The objective of this thesis is to develop a collective motion model based on the animal's natural visual perception. In this study, is assumed that each individual sees other individuals as spheres. Thus, individuals are not discretely identified, instead the recognition done only through projection on the panoramic retina of each individual. The visual information used by each individual is a combination of the optical size of one or a cluster of individuals, its retinal position, its apparent motion speed (translational optic flow), its optical expansion/contraction (optic flow divergence). With this visual information, the model is implemented using two rules followed by all the individuals, Attraction and Alignment.

The first rule, attraction, is based on the optical size of individuals as perceived by each individual in his visual field: this rule represents the group's inherent desire to stay together. Without attraction, individuals would naturally move away from each other.

The second rule, alignment, is based on optical flow. Note that this study makes a distinction between translational optic flow –corresponding to the optical scrolling of retinal objects– and optic flow divergence –corresponding to the optical size variation of retinal objects. Alignment reflects the group's tendency to move in the same direction. Without alignment, the individuals in the group would have difficulty keeping up with each other.

The proposed visual model bridges the gap between traditional particle physics-based approaches and visual perception. Moreover, this model can recreate the three main collective behaviors:

- swarming, like a group of bees or mosquitoes flying together in the same area,
- schooling, like a flock of birds flying together from one place to another, and,
- milling, like a multitude of fish circling in a swirling motion.

Furthermore, this visual model was successfully supplemented by introducing the necessary constraints for a robotized flock –delay, an anchor and avoidance rule–.

The results of the implementation on a flock of spherical robots closely mirrored those of the supplemented visual model. These similarities were not only observed on the behaviors that were reproduced, but also on their distribution in the alignment/attraction state space. The phase distribution map obtained is the first obtained with spherical robots, whose behavior is quite similar to that of traditional self-propelled particles.

Overall, my PhD thesis provides a biologically plausible framework for understanding collective motion. Accomplished with a constant parallel between bio-plausability and traditional self-propelled particles. And it's implementation of robots with similar kinematics.

Keywords: Collective movement, Visual Information, Bioinspired, Robotics, Vision, Optic-flow

Acknowledgments

Un día después del que sería tu cumpleaños terminamos lo que de la mano me llevaste a empezar cuando todo por conocer me faltaba. Me tomó diez años desde tu partida pero terminamos papá, siempre en mi corazón te llevo. A tu memoria dedico este trabajo.

Was not an easy road, and plenty I need to thank to get me to this point. But I can not start but with the beginning.

My dad, that life did not let him see me finish any university degree but believed in me as no one else, to his memory I have to thank. All those nights that he spend listening to me during the 2h drive home, after a 13 to 15h day, he was always as happy to hear my day no matter how insignificant it was. That bright curiosity that he fueled with his questions, with his patience, with his love. During the hardest challenges I faced, his memory gave me the strength I needed.

To my mom, that stood by me on the worse days, and her love filled the cup that was running low. With her daily calls that made the ocean that separate us fell just like a pond. Her meals that always arrived when she knew I had forgotten to eat. Her gentle, and no so gentle, reminders to keep my mind and body at check while I worked ignoring them.

To my tia Janeth, the second mom that life gifted me. Her care, her support and constant love despise she going through tough times. To my tio Emilio that always remember me despise the distance and a call was always on the right time.

To Yesid, thank you for your support when I needed it the most. For those 2 hour long calls where we discussed and manage to explain things no one else could understand. Thank you for been a role model, in a good way, thank you for being a terrific friend.

To Willy, you showed me the ropes and saw a bright future. Thank you for being so incredible and nurturing something that otherwise would have not flourish in that environment.

To Juli, life threw us together and it has been better ever since. You talks and lavadero, always made the work feel better afterwards. You care, your being there when I needed. Thank you.

Lastly Dani, that was there all the time, all the nights, all the days by my side, or on a facetime call. She endured the distance so I could finish something she did not asked for but she supported my dreams so we could reach the end. The tears, the missed sleep, the arguments you endured got us to the finish line, I can not express how much that helped. Thank you. C'est possible, C'est vraiment possible.

To my defense committee, thank you for your time, for attending my defense and reading 5 years worth of work condensed in 115 pages. I bet it was not easy, but thank you. I am delighted that you all could be here for this final step.

To Christophe Eloy, for giving me the chance to try to work on something incredible. For the valuable time that he always gave me, and the advice needed to keep on the right path. It was a truly life changing working with you. Thank you. I know I did not make it easy, but you made it memorable.

Now I am of those that believe that the start is as important as the finish, I had a wonderful start with my dad. But I could not have dream of a better finish than meeting Franck. Thank you Franck, I know it was not easy, I know it was challenging, I know the towel was temping, but we did it. During these past years I loved out 15-minute conversation that never were of 15 minutes, more on the 60 mark. Your constant guidance, patience and support made it possible, and at THAT time I really thought it was not. Thank you. Merci. Gracias. Arigato. Danke. I can not thank you enough.

Finally to you, thank you. I do not know who "you" would be, but thank you for taking your time and reading this work. It means a lot that these letters are being read by more than just those that had to read them and I really hope it helps you in the path that you are going.

Warm regards Diego

Contents

Affidavit 2

List of Publications 3

Résumé et mots clés 4

Abstract and keywords 6

Acknowledgments 8

Contents 10

List of Figures 13

List of Tables 17

List of Videos 18

1. Introduction 20
- 1.1. Animal Collective Motion . . . 21
- 1.2. Observing Nature . . . 24
  - 1.2.1. Microscale . . . 26
  - 1.2.2. Insects . . . 27
  - 1.2.3. Fish schools . . . 33
  - 1.2.4. Bird flocks . . . 36
  - 1.2.5. Lessons of observing nature . . . 38
- 1.3. Modeling Nature . . . 40
  - 1.3.1. Self-Propelled Particle Models . . . 40
  - 1.3.2. 3A models . . . 43
  - 1.3.3. Non-aligning models . . . 48
  - 1.3.4. Graph models . . . 49
- 1.4. Robotic collective motion . . . 51
- 1.5. Vision and animals . . . 54
  - 1.5.1. Optic flow . . . 58
  - 1.5.2. Measuring optic flow . . . 62
- 1.6. Vision and Collective motion . . . 64
- 1.7. Summary . . . 66

2. Reconciling vision and particle physics. 68
2.1. Abstract . . . 69
2.2. Introduction . . . 69
2.3. Model . . . 72
2.4. Numerical Simulations . . . 74
2.5. Discussion . . . 76
2.6. Conclusion . . . 77
2.7. Supplementary Information . . . 78
2.7.1. Mathematical development . . . 78
2.7.2. Digital implementation . . . 80
2.8. Reconciling vision and particle physics . . . 84

3. Robotic self-propelled particles. 88
3.1. Abstract . . . 89
3.2. Introduction . . . 89
3.2.1. Related Work . . . 89
3.2.2. Contribution . . . 91
3.3. Methods . . . 91
3.3.1. Robotic Implementation Challenges . . . 91
3.3.2. Robotic Setup . . . 92
3.3.3. Anchoring the collective behavior . . . 95
3.3.4. Robotic Visual Flocking Model . . . 96
3.3.5. Robots-in-the-loop computer architecture . . . 98
3.3.6. Metrics . . . 100
3.4. Results . . . 101
3.4.1. Simulation of Supplemented Visual Model . . . 101
3.4.2. Phase Diagrams of the Supplemented Model in Simulation . . . 102
3.4.3. Comparison with the phase diagram of the Spheros' robots-in-the-loop . . . 103
3.5. Discussion . . . 103
3.6. Conclusion . . . 106
3.7. Supplementary Information . . . 109
3.7.1. Spheros' dynamical response and time delay τ estimation . . . 109
3.7.2. Direct comparison between Supplemented visual model and robot-in-the-loop Experiments . . . 110

4. Conclusion 111
4.1. Summary . . . 111
4.2. Contribution . . . 117
4.3. Future Work . . . 118

ANNEXES 134

A. Supplementary information (Extra Results): Modeling collective behaviors from optic flow and retinal cues. 135

A.1. Extra Results . . . 135
A.1.1. Visual Model . . . 137
A.1.2. Omniscient Model . . . 139
A.2. Video Snapshots . . . 140
A.2.1. Time Step Analysis . . . 140
A.3. Phases Reproduction $k_\eta = 0.01$ Flock Size=50 . . . 153
A.3.1. Visual . . . 153
A.3.2. Omniscient . . . 156

B. Supplementary information: Visual collective behaviors on spherical robots 160

B.1. Image Processing Technical information . . . 160
B.1.1. Feature Matching Optic Flow . . . 161
B.2. Digital implementation of the supplemented visual model . . . 161
B.3. Observed phases . . . 164
B.4. Occlusion Analysis . . . 164

List of Figures

1.1. Examples of flocking. . . . 22
1.2. Examples of serpentine motion on droplets of water. . . . 23
1.3. Stereometry camera system. . . . 25
1.4. Example of votex-like behavior on bacteria. . . . 26
1.5. Example of votex-like behavior on bacteria. . . . 27
1.6. Example of swarming behavior on mosquitos. . . . 29
1.7. Example of swarming behavior on plant-animal worms. . . . 30
1.8. Example of collective motion in insects. . . . 32
1.9. Example of collective motion on fish. . . . 34
1.10. Fish predatory escape reaction. . . . 35
1.11. Bird flock flight formations. . . . 36
1.12. Birds collective motion example. . . . 37
1.13. Multiple species collective motion. . . . 39
1.14. *Boids* Model behavioral rules. . . . 41
1.15. *Vicsek* Collective motion reproduced. . . . 42
1.16. Zones of interactions. . . . 43
1.17. Main collective motion phases. . . . 45
1.18. Effect of zones of interactions on collective behavior reproduced. . . . 45
1.19. Model based on neighbors. . . . 46
1.20. Phase diagram and phases reproduced on Calovi et al. 2014. . . . 47
1.21. Phases reproduced on Strömbom 2011. . . . 48
1.22. Phases reproduced on Grossman et al. (2008). . . . 49
1.23. Collective motion as a graph-model. . . . 50
1.24. Some robotic platforms for collective motion. . . . 51
1.25. Behavioral collective motion on robots. . . . 52
1.26. 3A model in robots. . . . 53
1.27. Graph networks and robots. . . . 54
1.28. Compound eye of insects. . . . 55
1.29. Visual resolution of birds eye's. . . . 57
1.30. Fish illumination collective response. . . . 58
1.31. Optic Flow examples. . . . 59
1.32. Mercator projections of the optic flow vector field. . . . 60
1.33. Aperture problem and correspondence problem. . . . 61
1.34. Feature matching between time steps. . . . 63
1.35. Visualization neighbor selection strategies Schilling et al. 2022. . . . 64
1.36. Phases reproduced by Bastien et al. 2020. . . . 65

2.1. Visual Model, notation and visual cues examples 71
2.2. Illustration of the phases observed for $N = 50$ individuals 74
2.3. Phase diagrams for $N = 50$ individuals 75
2.4. Example of the bistability phase. 77
2.5. Effect of a blind spot. 80
2.6. Geometrical equivalent of the optic flow calculation. 81
2.7. Example of two consecutive visual fields. 83
2.8. Visual cue for attraction behavior 85
2.9. Phase diagram N=50, different visual attractions 86
2.10. Optic location seen by an individual during a stable milling phase. . . . 87

3.1. Modeling collective behaviors from optic flow and retinal cues Castro et al. 2024b: a) View of the flock positions. b) Representation of flock positions after recognition at time step t. Each i-particle is described by its XY-coordinates and heading unit vector $\mathbf{e}_i$, with heading angle θ_i. c) Point-of-view of particle *0*, showing the relative retinal positions of each particle. The arrows represent the perceived optic flow, with the white arrow indicating the apparent velocity ($\mathbf{V}_{i,k}$). Only the azimuthal component of $\mathbf{V}_{i,k}$ is shown for retinal object k as perceived by $i = 0$. d) $\mathcal{V}(\phi)$: Binary panorama representation of the field of view of particle *0*. e) Apparent distance $\mathscr{R}_i(\phi)$ for particle *0*. f) Optic flow $\mathscr{O}_i(\phi)$ perceived by particle *1*. g) Optic flow divergence $\mathscr{D}_i(\phi)$ perceived by particle *1*. . 92
3.2. Spherical robotic platform. a) Sphero Bolt ® wireless educational robots. b) Example of the robots' movement, c) Heading control by differential rotation on the two opposite inner wheels, d) Induced nose-up pitch during forward displacement. 93
3.3. Experimental Setup and external infrastructure. a) Photograph of the experimental setup. b) Sample image captured by the camera (Daheng Imaging MER2-230-168U3C camera equipped with a 4 mm C-mount lens). c) Whole-robot body image segmentation: Segmentation of the robot's body based on the captured image. d) Recognition of individual light sources and heading: Identification of the front and rear LEDs allows calculation of the robot's heading vector. e) Robot identification and virtualization: Assignment of a unique identifier to each robot and visualization of their positions and orientations. 94
3.4. Binary Panorama Reconstruction. a) 2D virtual representation of the binary panorama: The red goal indicates the desired location of the flock. b) Mixed-reality perspective from particle 0: Displays the omniscient red goal as observed from the particle's point of view. c) Graph representation of all $\mathcal{V}(\phi, t)$: $\mathcal{V}(\phi, t)$ represents the binary panorama view of the field of vision for particle i, with the anchor displayed as a distinct channel (red). This graph can be efficiently stored as a tuple list, representing the rising and falling edges of the binary panorama. . . . 96

3.5. Functional diagram of the collective motion algorithm with robots-in-the-loop. 99
3.6. General Block Diagram. The N flocking robots are observed using a 1920x1200 px, 164 fps camera. The captured images are processed to determine the XY-coordinates, heading, and colors of all N robots. Based on these coordinates, each robot is uniquely identified. The XY-coordinates of the Anchor can be set at any point within the image to assist in robot identification. These coordinates are further processed to reconstruct the binary panorama for each spherical robot. Each binary panorama serves as input to an independent collective motion model, which processes the visual panoramic data. These models operate in parallel, and their outputs, in terms of heading, are transmitted to each robot via a Bluetooth interface. 99
3.7. Illustration of the phases observed in simulation for the supplemented visual model with the addition of avoidance, anchor, and time delay as defined in Eqs. 3.2, with $\tau = 2.4[a/U]$. Three phases are observed when varying the other two parameters: a) Swarming ($k_\odot = 0.25$, $k_\parallel = 0.04$) (see SVideo 1); b) Bistable ($k_\odot = 0.2$, $k_\parallel = 0.25$) (see SVideo 2); c) Milling ($k_\odot = 0.07$, $k_\parallel = 0.16$) (see SVideo 3). 101
3.8. Simulation Phase Diagrams for 10 individuals. The dotted line indicates $P = 0.5$, the dashed line indicates $M = 0.5$, and the solid line indicates $O = 0.3$. a) Color code, b) Original visual model Castro et al. 2024b, c) Supplemented visual model with avoidance term, $\mathscr{A}(\mathscr{R}(\phi,t))$, d) Supplemented model with avoidance and anchor term, $\omega_c(t)$, e) Supplemented visual model with avoidance, anchoring, and perceived time delay ($\tau = 2.4[a/U]$). 102
3.9. Experimental results for a flock of 10 independently controlled robots (with $U = 4a/sec = 0.14[m/sec]$ (the radius of Spheros robot is $a = 3.5cm$). a) Phase diagram, the dotted line is $P = 0.5$, the dashed line is $M = 0.5$, and the solid line is $O = 0.7$. b) Swarming ($k_\odot = 0.25$, $k_\parallel = 0.04$) (see raw footage SVideo 4, robotic schematic SVideo 5 or a 3rd person view SVideo 6); c) Bistable ($k_\odot = 0.2$, $k_\parallel = 0.25$) (see raw footage SVideo 7, robotic schematic SVideo 8 or a 3rd person view SVideo 9); d) Milling ($k_\odot = 0.07$, $k_\parallel = 0.16$) (see raw footage SVideo 10, robotic schematic SVideo 11 or a 3rd person view SVideo 12). 104
3.10. Spheros' dynamical response 109
3.11. Difference between the Phase Diagrams. 110

4.1. Expanded Phase diagram of the supplemented model. 116

A.2. Visual Phase Diagram and Phase Reproduction for different Flock sizes 137
A.3. Phase diagram for the Visual model N=50, $\eta = 0.01$ and $\epsilon = 0$ 138
A.4. Phase diagram for the Visual model N=50, $\eta = 0.01$ and $\epsilon = 1$ 138
A.5. Phase diagram for the Visual model N=50, $\eta = 0.1$ and $\epsilon = 0$ 139

A.6. Phase diagram for the Visual model N=50, $\eta = 0.1$ and $\epsilon = 1$ 139
A.1. Phase diagrams for $N = 50$ individuals. 141
A.7. Observed phases of omniscient model N=50 142
A.8. Phase Diagram of the omniscient model 143
A.9. Phase Diagram of the omniscient model with several flock sizes 144
A.10.Phase diagram for the Omniscient model N=50, $\eta = 0.01$ and $\epsilon = 0$ . . . 145
A.11.Phase diagram for the Omniscient model N=50, $\eta = 0.01$ and $\epsilon = 1$ . . . 145
A.12.Phase diagram for the Omniscient model N=50, $\eta = 0.1$ and $\epsilon = 0$ 146
A.13.Phase diagram for the Omniscient model N=50, $\eta = 0.1$ and $\epsilon = 1$ 146
A.14.Video Snapshot for Visual model δt=0.001 $k_\odot$ =0.06,$k_\parallel$ =0.06 and k_η =0.01. 147
A.15.Video Snapshot for Visual model δt=0.01 $k_\odot$ =0.06,$k_\parallel$ =0.06 and k_η =0.01. 148
A.16.Video Snapshot for Visual model δt=0.1 $k_\odot$ =0.06,$k_\parallel$ =0.06 and k_η =0.01. .149
A.17.Video Snapshot for Omniscient model δt=0.001 $k_\odot$ =0.06,$k_\parallel$ =0.06 and k_η =0.01. 150
A.18.Video Snapshot for Omniscient model δt=0.01 $k_\odot$ =0.06,$k_\parallel$ =0.06 and k_η =0.01. 151
A.19.Video Snapshot for Omniscient model δt=0.1 $k_\odot$ =0.06,$k_\parallel$ =0.06 and k_η =0.01. 152
A.20.Visual Swarming N= 50, k_η =0.06 k_η =0.0 k_η =0.01 and no blind angle . 153
A.21.Visual Schooling N= 50, k_η =0.06 k_η =0.2 k_η =0.01 and maximum blind angle 154
A.22.Visual Milling N= 50, k_η =0.06 k_η =0.06 k_η =0.01 and maximum blind angle 155
A.23.Omniscient Swarming N= 50, k_η =0.06 k_η =0.0 k_η =0.01 and no blind angle 157
A.24.Omniscient Schooling N= 50, k_η =0.06 k_η =0.2 k_η =0.01 and maximum blind angle 158
A.25.Omniscient Milling N= 50, k_η =0.06 k_η =0.06 k_η =0.01 and maximum blind angle 159

B.1. Image Processing. 161
B.2. Optic flow example. 162
B.3. Retinal object matching between consecutive time steps. 163
B.4. Observed phases original model Castro et al. 2024b 164
B.5. Observed phases original model with avoidance 165
B.6. Observed phases original model with the avoidance and visual anchor 165
B.7. Number of collisions 166
B.8. Minimum inter-distance 166

List of Tables

3.1. *Daheng Imaging MER2-230-168U3C* global shutter Camera Settings. . 95

B.1. Camera: *Daheng Imaging MER2-230-168U3C* specifications. 160

List of Videos

2.1. Visual model: Swarming phase 74
2.2. Visual model: Schooling phase 74
2.3. Visual model: Milling phase 74
2.4. Visual model: Bistable State 77

3.1. Simulation Swarming phase (SVideo 1) 101
3.2. Simulation Bistable state (SVideo 2) 101
3.3. Simulation Milling phase (SVideo 3) 101
3.4. Robotic Swarming phase: Raw Data (SVideo 4) 104
3.5. Robotic Swarming phase: Robot Schematic (SVideo 5) 104
3.6. Robotic Swarming phase: 3rd person view (SVideo 6) 104
3.7. Robotic Bistable State: Raw Data (SVideo 7) 104
3.8. Robotic Bistable State: Robot Schematic (SVideo 8) 104
3.9. Robotic Bistable State: 3rd person view (SVideo 9) 104
3.10. Robotic Milling phase: Raw Data (SVideo 10) 104
3.11. Robotic Milling phase: Robot Schematic (SVideo 11) 104
3.12. Robotic Milling phase: 3rd person view (SVideo 12) 104

A.1. Video Snapshot for Visual model δt=0.001 $k_{\odot}$ =0.06,$k_{\parallel}$ =0.06 and k_{η} =0.01. 147
A.2. Video Snapshot for Visual model δt=0.01 $k_{\odot}$ =0.06,$k_{\parallel}$ =0.06 and k_{η} =0.01. 148
A.3. Video Snapshot for Visual model δt=0.1 $k_{\odot}$ =0.06,$k_{\parallel}$ =0.06 and k_{η} =0.01. .149
A.4. Video Snapshot for Omniscient model δt=0.001 $k_{\odot}$ =0.06,$k_{\parallel}$ =0.06 and k_{η} =0.01. 150
A.5. Video Snapshot for Omniscient model δt=0.01 $k_{\odot}$ =0.06,$k_{\parallel}$ =0.06 and k_{η} =0.01. 151
A.6. Video Snapshot for Omniscient model δt=0.1 $k_{\odot}$ =0.06,$k_{\parallel}$ =0.06 and k_{η} =0.01. 152
A.7. Swarming phase reproduction for a flock size of 50 individuals for the Visual model. k_{η} =0.06 k_{η} =0.0 k_{η} =0.01 and no blind angle 153
A.8. Schooling phase reproduction for a flock size of 50 individuals for the Visual model. k_{η} =0.06 k_{η} =0.2 k_{η} =0.01 and maximum blind angle . . . 154
A.9. Milling phase reproduction for a flock size of 50 individuals for the Visual model. k_{η} =0.06 k_{η} =0.06 k_{η} =0.01 and maximum blind angle 155
A.10. Swarming phase reproduction for a flock size of 50 individuals for the Omniscient model. k_{η} =0.06 k_{η} =0.0 k_{η} =0.01 and no blind angle 157

A.11.Omniscient Schooling N= 50, k_η =0.06 k_η =0.2 k_η =0.01 and maximum blind angle 158
A.12.Omniscient Milling N= 50, k_η =0.06 k_η =0.06 k_η =0.01 and maximum blind angle 159

B.1. Original model: Swarming phase 164
B.2. Original model: Bistable phase 164
B.3. Original model: Milling phase 164
B.4. Original model with avoidance: Swarming phase 165
B.5. Original model with avoidance: Bistable phase 165
B.6. Original model with avoidance: Milling phase 165
B.7. Original model with avoidance and visual anchor: Swarming phase . . 165
B.8. Original model with avoidance and visual anchor: Bistable phase . . . 165
B.9. Original model with avoidance and visual anchor: Milling phase 165

1. Introduction

Contents

1.1. Animal Collective Motion 21
1.2. Observing Nature 24
1.2.1. Microscale 26
1.2.2. Insects 27
1.2.3. Fish schools 33
1.2.4. Bird flocks 36
1.2.5. Lessons of observing nature 38
1.3. Modeling Nature 40
1.3.1. Self-Propelled Particle Models 40
1.3.2. 3A models 43
1.3.3. Non-aligning models 48
1.3.4. Graph models 49
1.4. Robotic collective motion 51
1.5. Vision and animals 54
1.5.1. Optic flow 58
1.5.2. Measuring optic flow 62
1.6. Vision and Collective motion 64
1.7. Summary 66

"*More and more researchers are interested in this new exiting way of achieving a form of artificial intelligence, swarm intelligence– the emergent collective intelligence of group of simple agents.*"(Bonabeau et al. 1999).

A statement that was made at the start of the millennia, and is as valid as it did back then. After almost 30 years of research, we are yet to grasp the complete inner workings of this emergent collective intelligence. Even more, we have yet to explain the simpler phenomena of the emerging patterns of collective motion, when these are ubiquitous to most animal groups. Meaning that no matter if it is a swarm of insects, a colony of bacteria, a school of fish, a flock of birds or, even, mammal herds; they all display a form of collective motion or collective behavior (Vicsek et al. 2012). These patterns are displayed regardless of the different sizes, habitat, locomotion, perception capabilities or social interactions that each animal group is known for (Calovi et al. 2014). Imagine the capabilities of a collective or cooperative artificial system that would work towards the same goal and with the same degree of flexibility. True collective intelligence.

The first steps, towards reaching this collective intelligence, are to understand, model and replicate the aforementioned emerging patterns. And even though there are truly impactful works such as Couzin et al. (2002), Vicsek et al. (1995), Bouffanais (2016), Theraulaz et al. (2003) or Morin et al. (2015); so far, our advancements in this basic collective intelligence lacks the connection with the animal's ability to perceive its environment. In other words, our current understanding of natural phenomena is build upon artificial concepts, that do not have a parallel in nature.

This work aims at breaching that gap, to simulate and replicate nature from proven concepts that are found, and used, in nature. Without losing, introducing new or complex systems that have no parallel in the literature.

1.1. Animal Collective Motion

But then what is Collective motion?. Collective motion refers to the synchronized movement of individuals within a group, observed in a wide range of systems, both biological and non-biological (Vicsek et al. 2012). Collective motion is the natural occurring behavior that makes an individual's action be mainly caused by the influence of "others"(Bazazi et al. 2011). Another definition could be that is a phenomenon that makes individuals coordinate their actions without centralized control. In summary, collective motion is generally described as phenomena where a group heterogeneously modify their behavior to a common pattern (Papadopoulou et al. 2022). Furthermore, the same "pattern" can be seen across vastly different biological conditions, from different length scales- microscopic bacteria colonies as seen on H.-P. Zhang et al. (2010) to large mammal herds (Sinclair 1977). The pattern is also invariant to different types of locomotion, for example Krongauz et al. (2024) showcases crawling, Ballerini et al. (2008a) show flying or even swimming as seen on Bonnen (2023). Some examples of nature collective motion can be seen on Fig. 1.1.

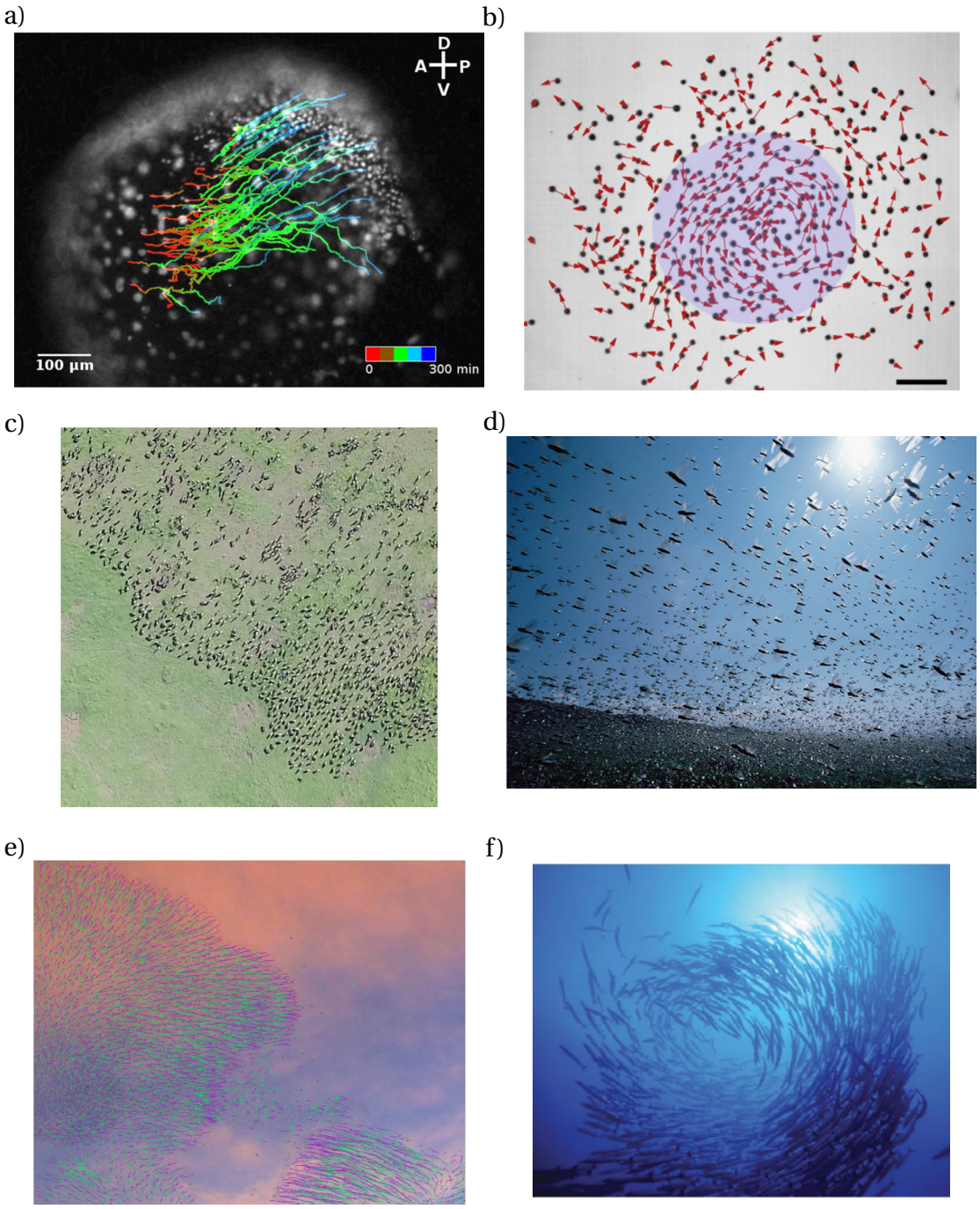

Figure 1.1.: Examples of flocking. a) Collective cell motion during embryonic zebrafish development (reprinted from Battersby 2015), b) Ferromagnetic particles black scale 1mm reprint from Kaiser et al. 2017, c) Wildebeest herds (reprinted from Hughey et al. 2018), d) Locust Migration (reprinted from Ren et al. 2016), e) Flock of starlings, 2D trajectories of individual birds. The green color represent earlier positions while purple presents the current ones (reprinted from Moroni et al. 2008), f) School of fish (reprint from Lopez et al. 2012).

Lin et al. (2024) have extended the concept of collective motion to non-living systems. They have demonstrated that self-propelled condensate droplets on thin films move in coordinated serpentine and circular patterns (Fig. 1.2), revealing that these collective patterns are not only found on living systems.

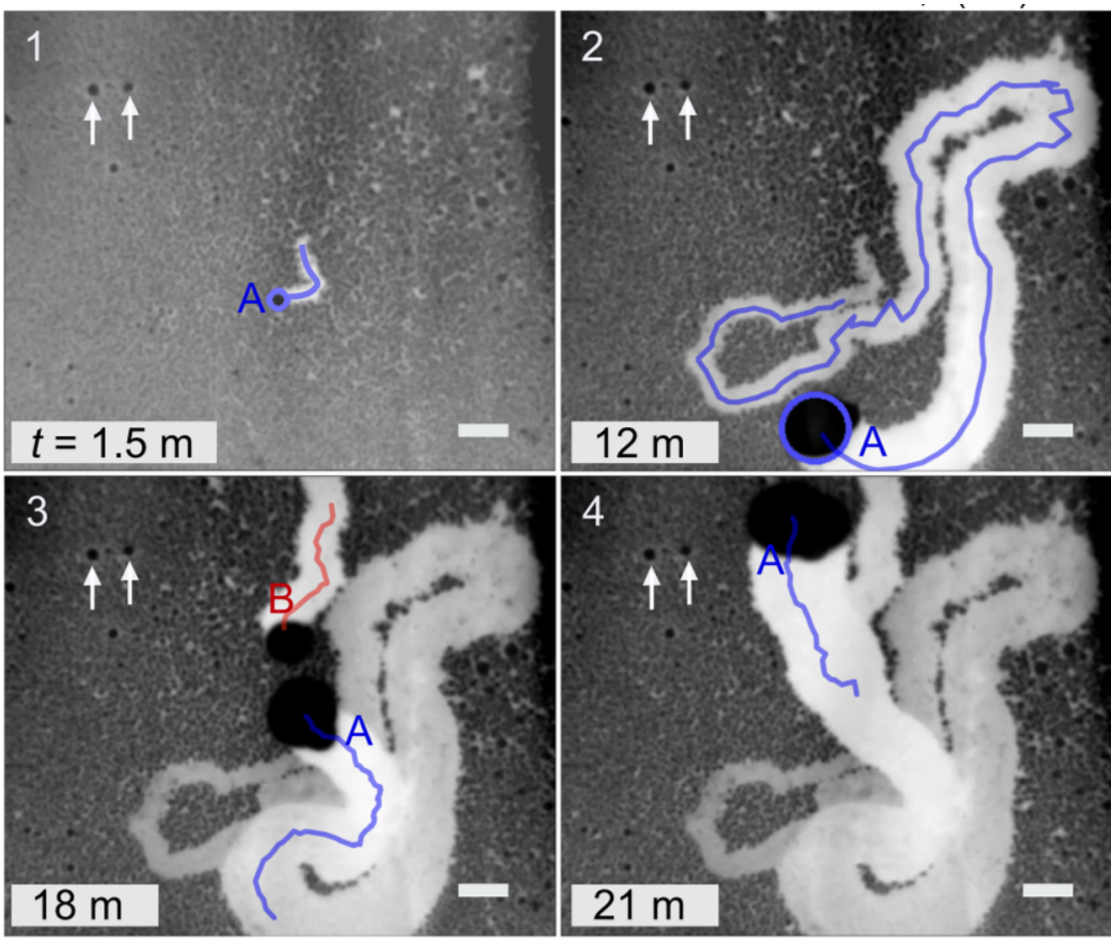

Figure 1.2.: Examples of serpentine motion on droplets of water (reprinted from Lin et al. 2024(.

The collective patterns are not only striking to observe but also provide key benefits to the group (Attanasi et al. 2014; Nagy et al. 2013; Katz et al. 2011; Wey et al. 2008; Buhl et al. 2006; Budrene et al. 1991; Papadopoulou et al. 2022; Bonabeau et al. 1997). This key benefits include those presented by Caraco et al. (1980), Papadopoulou et al. (2023a), and Procaccini et al. (2011), showing how collective motion provides better protection from predators. Caraco (1981), Baker et al. (1981), Fernández-Juricic et al. (2004), and Aplin et al. (2014) show documented cases of collective motion enhancing the efficiency of foraging for food and water. Gökçe (2008), Cavagna et al. (2013), and Sumpter et al. (2008) are examples of studies that proved that collective motion reduce the energetic costs of long-distance. Not only the benefits to the group are clear but collective patterns demonstrate a broad presence in the animal world. Furthermore,

this universality hints that the governing underlying principles of collective motion are independent to the specific environtment, size or defining characteristics of the animals or agents involved.

The following sections will present some examples of natural occurring collective motion on different time and length scales, present current theories to explain collective motion from a natural perspective.

1.2. Observing Nature

To understand the collective motion phenomena, we must bring forth the layers that construct these fascinating systems. And eventually we are able to use these new understandings to develop science and technology. The first challenge to overcome was the identification and recording of each single individual belonging to the same flock. This obstacle can be particularly deceptive in its difficulty. As a wild, flock can easily be thousands of similarly looking individuals. In other words, the same natural advantage that a flock has over predators, presents a challenge for data gathering. An individual becomes hard to detect and/or follow while in the flock. Furthermore, the scales from which flocking appears makes a single data gathering technique not be possible.

Collective motion patterns are seeing on microscale systems, such as the single cellular flocks presented by Kudrolli et al. (2008) or particle-like shown by Das et al. (2024). Then some specialized techniques had to be developed to see the movement of these microscopic systems. The most common techniques are based on particle image velocimetry (PIV). PIV works by seeding fluid flows that carry minuscule particles that act as tracers, that allows to perceive the motion pertaining to the tracer particles for derivation of fluid velocities (Etminan et al. 2022). Measuring PIV requires illuminating the area (usually a laser), a recording device (photographic film or digital cameras) at defined points in space or time. The scattering of the passing laser light allows seeing the movement of the desired particles and have the traceability desired for this tiny scale. This technique is most suitable for collective motion analysis, as it presents:

- Non-intrusive velocity measurement.
- Indirect velocity measurement.
- High resolution.
- Challenge on recognition.
- Dependent on initial seeding and noise to tracer particle ration.

Tracking the movement of bigger individuals has largely relied on camera-based techniques. Sinclair (1977) presented the earliest recorded efforts in documenting large flocks. Specifically, it presented a recording studying buffalo herds using aerial photos. The aerial photos of herds presented a two-dimensional (2D) plane of the flock

movement. Three-dimensional (3D) tracking, on the other hand, presented a greater challenge. Cullen et al. (1965), introduced the *shadow method* for fish, as its name implies uses fish projection's (Similar to a shadow) to better track their movements. Using this technique, Parrish et al. (1997) were able to capture fish in 3D by using a set of three synchronized cameras (one per axis). Becco et al. (2006) introduced another set of techniques for confined fish in a shallow container. The reduced depth of the container reduced the dimensionality of the problem to be bound to the plane. The "shadow" and "shallow depth" methods presented great success and is still used today for fish recording, although they are not suited for birds or terrestrial animals. To overcome that issue, Major et al. (1978) used stereo photography to capture bird flock positions. A technique that today has been successfully used to capture hundreds of starlings (Ballerini et al. 2008b). The general concept of the method can be seen in Fig. 1.3. Using two or more cameras, separated with a known distance and an angle, allows having depth information. Together, the depth and location information allows performing a better recognition and tracking of a given individual.

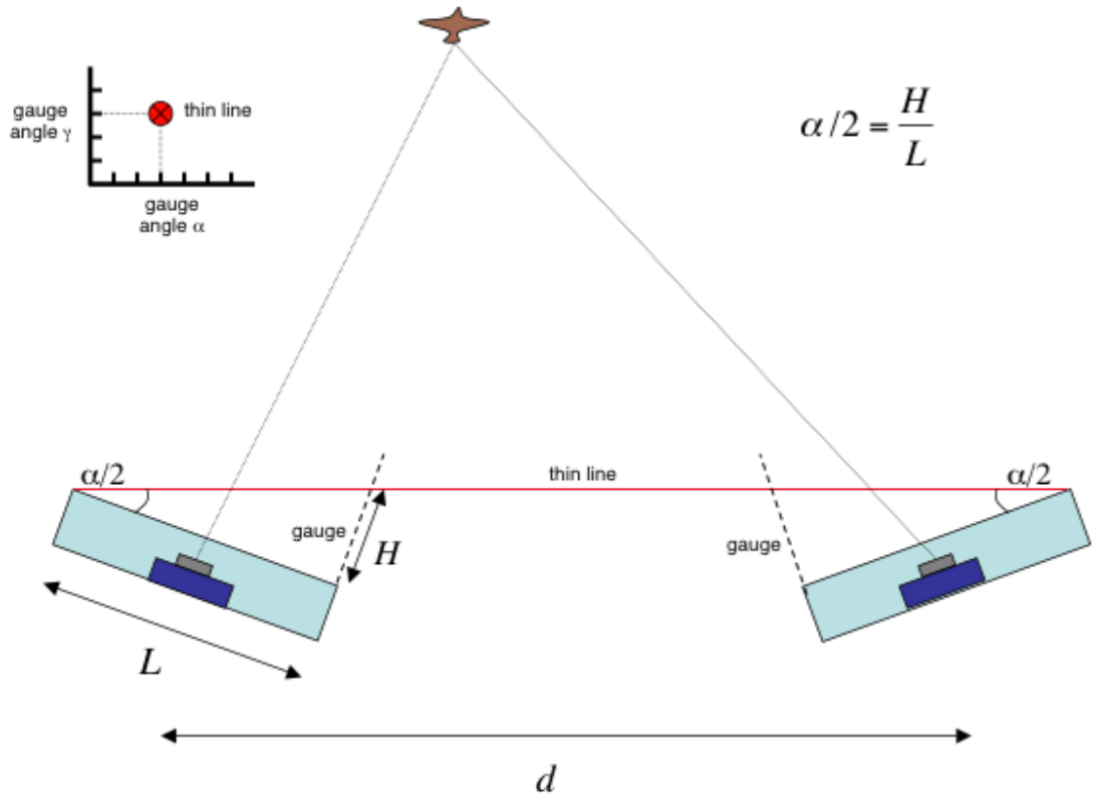

Figure 1.3.: Example of stereometry camera system for Starlings (reprinted from Cavagna et al. 2008a).

Regardless of the camera-based method employed, most will face an issue on boundaries, resolution or moving with the flock. A more recent approach has been to equip

each individual with a small device capable of localization, either via Global Position System (GPS) or UltraWideBand (UWB). Zaitouny et al. (2017) successfully tracked a small flock of pigeons over large distances. But this approach presents several disadvantages. The external nature of this kind of device presents high complexity to attach the device to the animal, even more without affecting the outcome of the study. Furthermore, it presents high implementation cost with low reliability, specially with tenths of hundreds connected devices on a small area.

1.2.1. Microscale

The interactions between individuals give birth to collective motion, therefore, is a phenomenon that is not dependent on the length-scale of the flock. Keller et al. (1971) presented the earliest example of microscale collective motion, particularly *Escherichia coli* bacterial colonies. The study developed a theoretical model using partial differential equations to describe bacterial density changes due to random motion and chemotaxis. András Czirók et al. (1996), presented a mathematical model of the vortex-like pattern on colonies of a morphotype of *Bacillus subtilis* (Fig. 1.4). The proposed model was based on a ferromagnetic analog, following the same formulation that Vicsek et al. (1995) had for collective motion. The model incorporated control parameters, such as the concentration of *agar* and *peptone*, which was the source of subtract of the Petri dish.

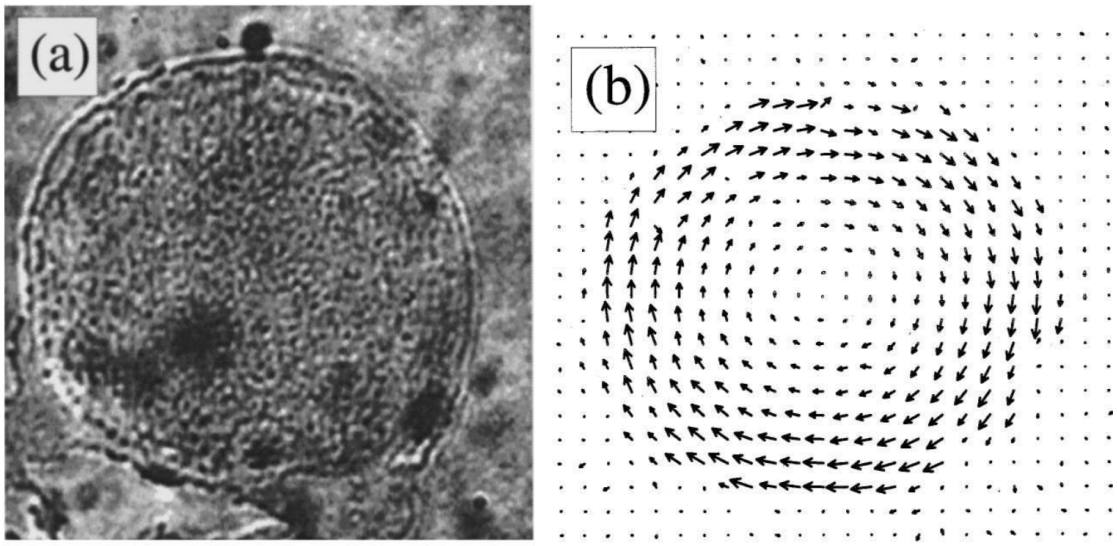

Figure 1.4.: Example of vortex-like behavior on a colony of morphotype of *Bacillus subtilis* (reprinted from András Czirók et al. 1996).

Andras Czirók et al. (2001) explored the swarming cycles observed in *Proteus mirabilis* colonies. These cycles result in a characteristic concentric ring structure as bacteria periodically shift between "vegetative" and "mobile" states. During each cycle, bacteria at the colony margin differentiate into mobile cells that migrate outward,

only to revert to vegetative cells through fission. Yamazaki et al. (2005), explored the transition between "vegetative" and "mobile" cell states in *Proteus colonies*. The transition is not governed by biological or chemical factors, but rather by local cell density. Furthermore, Lushi et al. (2014), presented a model that takes into account the hydrodynamic interaction between the cells and the medium in which they are suspended. These interactions affect long-range interaction between cells. Their model align with experimental studies that track the orientation of cells in a confined, dense bacterial suspension (Fig. 1.5).

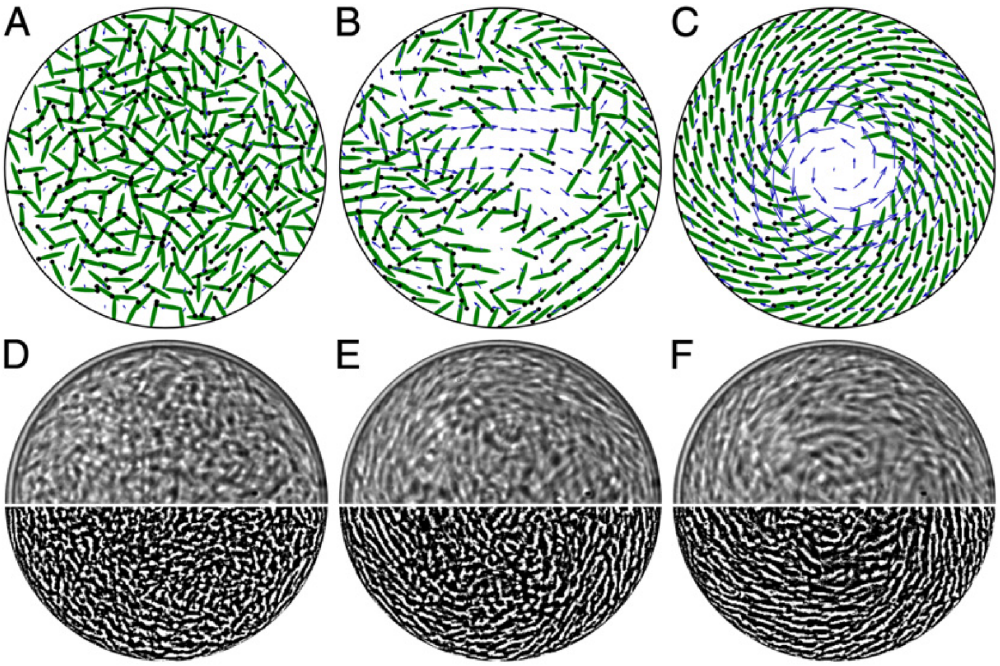

Figure 1.5.: Example of vortex-like behavior on a colony of morphotype of *Bacillus subtilis* (reprinted from Lushi et al. 2014).

These are a few examples of how on a microscopic scale, the organisms react to others and their environment and collective motion appears. Note that in microscale most information available to the organism is close range (chemotaxis, hydro dynamic interactions, touch), and yet patterns that highly benefit from long distance information appear.

1.2.2. Insects

Insects are the most abundant, diverse and varied animal groups on the planet (Chapman et al. 2009). Furthermore, some of the most social animals are insects (ants as seen on Gordon (1996), bees as seen on Page Jr et al. (2001)). A. M. Reynolds (2023), presented a system to document and evaluate in-lab swarming behavior of mosquitos (*Anopheles gambiae*). The study presented several gender differences in the flying

patterns of the individuals, where mating interactions lead to distinct behaviors between males and females. Furthermore, it was shown that the males tend to stay in the center of the flock while flying on a circular motion, while the females have a more swarm-like pattern.

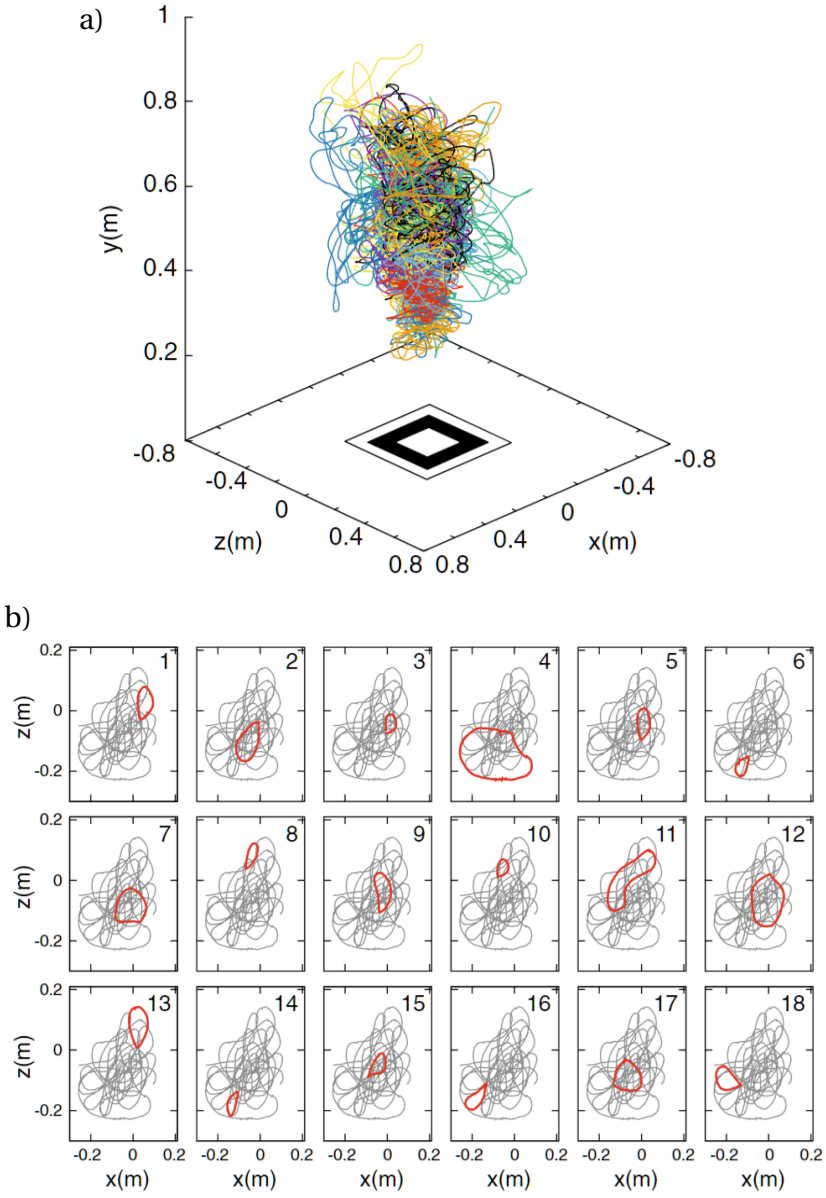

Figure 1.6.: Characteristic circular movements in the horizontal plane. (a) The 3D reconstructed trajectories of a single-sex male swarm. Different colors represent different mosquitoes, (b) The trajectory of a single individual in the horizontal plane, as it would be seen from above the swarm. Mosquitoes perform consecutive pseudocircular motions, highlighted in red, superimposed on the complete trajectory of the same individual, shown in light gray. Numbers in the top right corner of each sub-panel represent the temporal order of the ring-like movements (reprinted from A. M. Reynolds 2023).

Buhl et al. (2006) studied the effect of locust density on their transition from disordered to ordered states of movement, finding that coordinated marching behavior

became more pronounced as density increased. This transition aligns with predictions from a classical self-propelled particle (*SPP*) model, particularly the model presented by Vicsek et al. (1995). Using Vicsek's models as a base, a critical density threshold was identified for the onset of coordinated movement. Yates et al. (2009) followed the research on locust swarms, focusing on the sudden coherent changes in direction, using a self-propelled agent-based model. The model propose successfully reproduced the observed behavior, including short-range velocity correlations and an attraction to a central point.

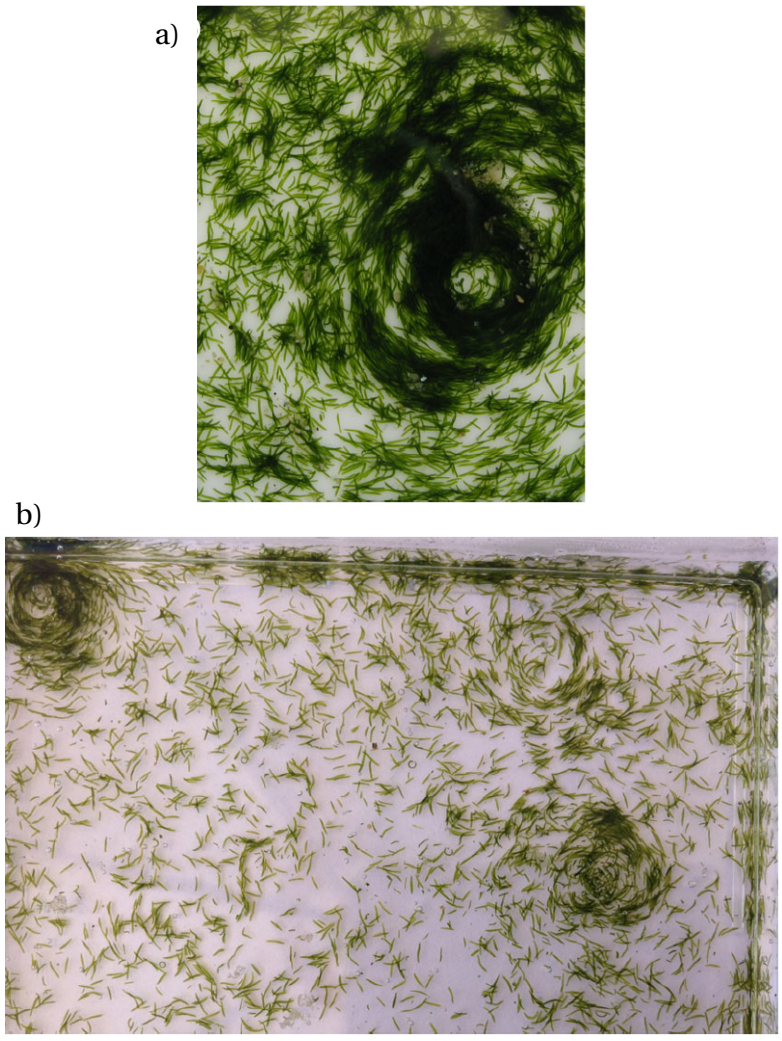

Figure 1.7.: Example of swarming behavior on plant-animal worms. *Symsagittifera roscoffensis* worms exhibit milling as a step function of increasing density. a) A circular mill in a small arena, b) Part of a 100 x 100 mm Petri dish where three circular mills have formed (reprinted from Franks et al. (2016)).

Franks et al. (2016) demonstrated the self-organizing social behaviors in the plant-animal worm, *Symsagittifera roscoffensis*. These worms have a symbiotic relationship

with an algae (hence the name). The worm coordinate their movements among neighbors so that at low densities, small polarized groups form. While at higher densities, the same group turn into circular mills (Fig. 1.7). This increase of density becomes detrimental to the worm's food source (the nutrients generated by the photosynthesis of its symbiotic algae), which presents an interesting case study for the paradox between self-desire and collective benefits.

Interestingly, not all collective behavior is positive. Bazazi et al. (2008) presented how juvenile desert locusts migration was influenced by cannibalistic interactions. In this study, it was seen that individuals in marching bands often bite the locust in front. By reducing the capacity to detect the attack of others from behind by surgical methods, they decreased the locust's probability to move. This dramatically reduced the amount of moving individuals in the group, which in turn increased cannibalism. Interestingly, their tendency to advance was not dominated by the abdominal biting or the approach of other from behind, but rather the rear visual field. Another case of cannibalism while in collective behavior is presented by Bazazi et al. (2010). They observed that Mormon crickets exhibit cannibalistic behavior triggered by both, the position of others and the social conditions. Crickets were more likely to attack side-located stationary individuals in crowed environments.

To conclude, insects are the most common animal on the world and there are several species that have shown collective behavior, some of them are:

- Monarch buterfly (swarming and schooling) (Reppert et al. 2010)
- Honeybees (swarming and schooling) (Vries et al. 1998)
- Locust (swarming and schooling) (Bazazi et al. 2008)
- Ants (swarming. milling and schooling) (Schneirla et al. 1948)
- Worm (swarming and milling) (Bazazi et al. 2010)
- Mormon cricket (swarming and schooling) (Sword 2005)
- Mosquito (swarming and schooling) (Franks et al. 2016)

a)



b)



c)

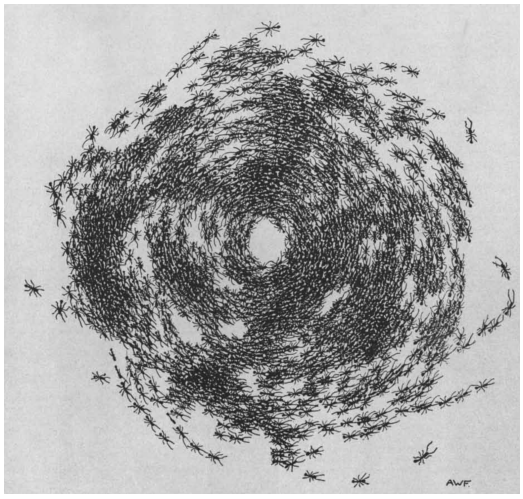

Figure 1.8.: Example of collective motion in insects. a) Ants swarming (taken from flickr under creative commons), b) Bees swarming (taken from flickr under creative commons), c) Ants milling (reprinted from Schneirla et al. 1948).

1.2.3. Fish schools

Aquatic animals are the groups that are most associated with collective motion. Both fish schools and shoal of fishes naturally present a variety of patterns (Fig. 1.9). Although, the most while spread pattern in aquatic species is schooling (Shaw 1978). Shaw (1962) pioneer the field by laying the groundwork of this main patter of collective motion on fish, schooling. They studied wild behaviors during most of the growing cycle of two species of *Menidia*, known commonly as whitebait. The schooling behavior did not appear until the fish were big enough to be recognized by Shaw's team. This recognition allowed the identification of the sensory tools —particularly vision and the lateral line— that allow fish to maintain synchronized movement. This behavior, schooling, was later reproduced on a controlled environment. This research established that schooling is a highly coordinated, internally driven behavior.

Partridge (1982) expanded on the highly coordinated behavior by introducing the probabilistic nature of school structure and highlighting the role of sensory systems in maintaining group cohesion. It was emphasized, that the adaptive value of schooling is not just for predator avoidance but also for increasing the chances of survival through confusion tactics (Fig. 1.10). Furthermore, it linked the peripheral vision with maintaining speed and positioning within schools. This was the first evidence that collective behavior is a result of complex, multisensory integration.

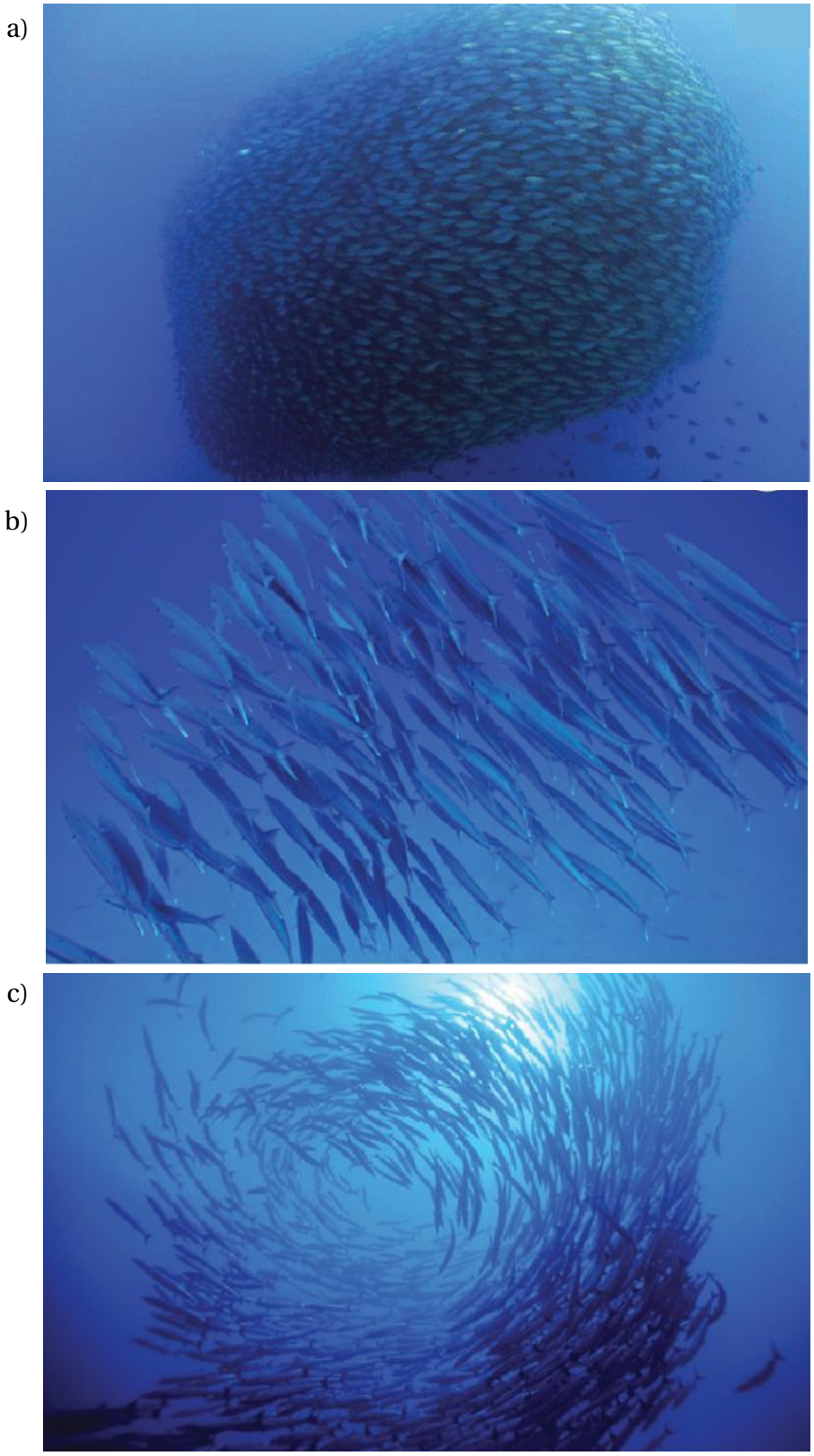

Figure 1.9.: Example of collective motion on fish. a) Swarming: condensed, non-orderly flock, b) Schooling: orderly, highly aligned motion, c) Milling: vortex-like motion (reprinted from Lopez et al. 2012).

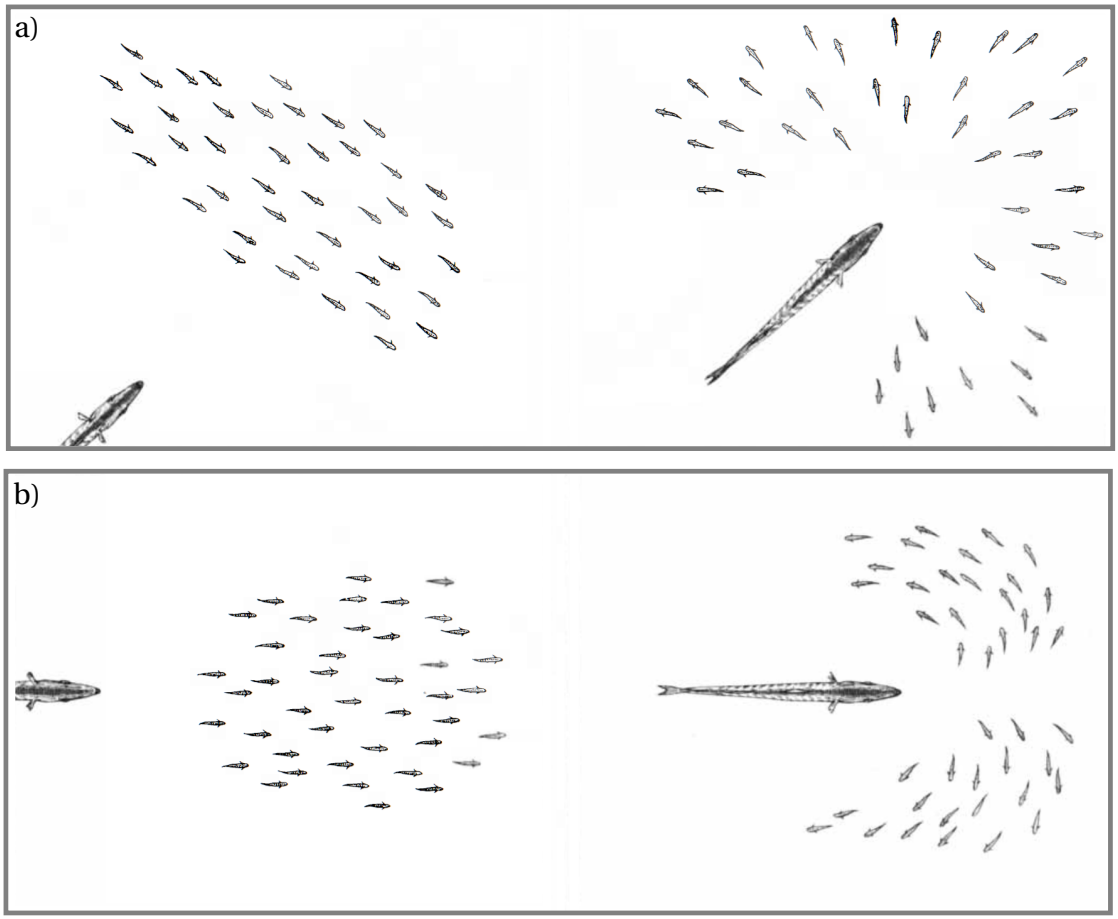

Figure 1.10.: Fish predatory escape reaction. a) Indiscriminate spreading of the flock when a predator tries to interact with the flock, b) Coordinate spreading of the flock when a predator tries to interact with the flock (reprinted from Partridge 1982).

Miller et al. (2012) performed experiments on groups of zebrafish (*Danio rerio*) that showed the first multi-pattern examples of collective motion. They found that the collective behaviors that zebrafish present, are bimodal. According to the level of stress, the school could change from schooling to swarming[1]. The study presented experiments on a controlled environment with different flock sizes from 5 to 50 individuals, suggesting that larger groups may swarm more frequently, especially under low-stress conditions. Schooling was presented as a more common response in smaller, stressed groups. Lafoux et al. (2024) also presented experimental collective motion on a controlled environment. Notably, they found that a group of rummy-nose tetra fish (*Hemigrammus rhodostomus*) presented two different collective motion patterns (milling and schooling) at the same time. Better known as a bistability state. A collective motion system is considered bistable when it oscillated between two

[1]Miller et al. 2012 referred to this state as shoal, but the definition of shoal presented previously would have presented confusion to the reader of this manuscript.

distinct phases. Lafoux et al. (2024) bistability was described as a Markov process.

Lopez et al. (2012) explores the principles of collective animal motion and develops models for understanding how simple individual rules can give rise to complex group behaviors such as schooling in fish. It emphasizes the role of local interactions among individuals in shaping the global behavior of the group. The study also discusses how factors like alignment, attraction, and repulsion influence collective dynamics and how these models can apply across species, offering insights into the adaptive value of coordinated group motion.

1.2.4. Bird flocks

Emlen (1952) is one of the earliest studies that tackles bird flocking from a social perspective is presented. From observations and similar behaving groups, they proposed that the patterns were caused by both positive and negative social forces on each individual. Heppner (1974) presented the formations seen by migratory birds with a detailed discussion about the advantages and disadvantages of these formations (Fig. 1.11). To explain these formations it was hinted that it should be something further than aerodynamics, maybe a possible visual communication. Weimerskirch et al. (2001) presented a link between the V-flight pattern of migrating geese, and showed a significant decrease of the heartbeat of peak heartbeat of white pelicans (*Pelecanus onocrotalus*).

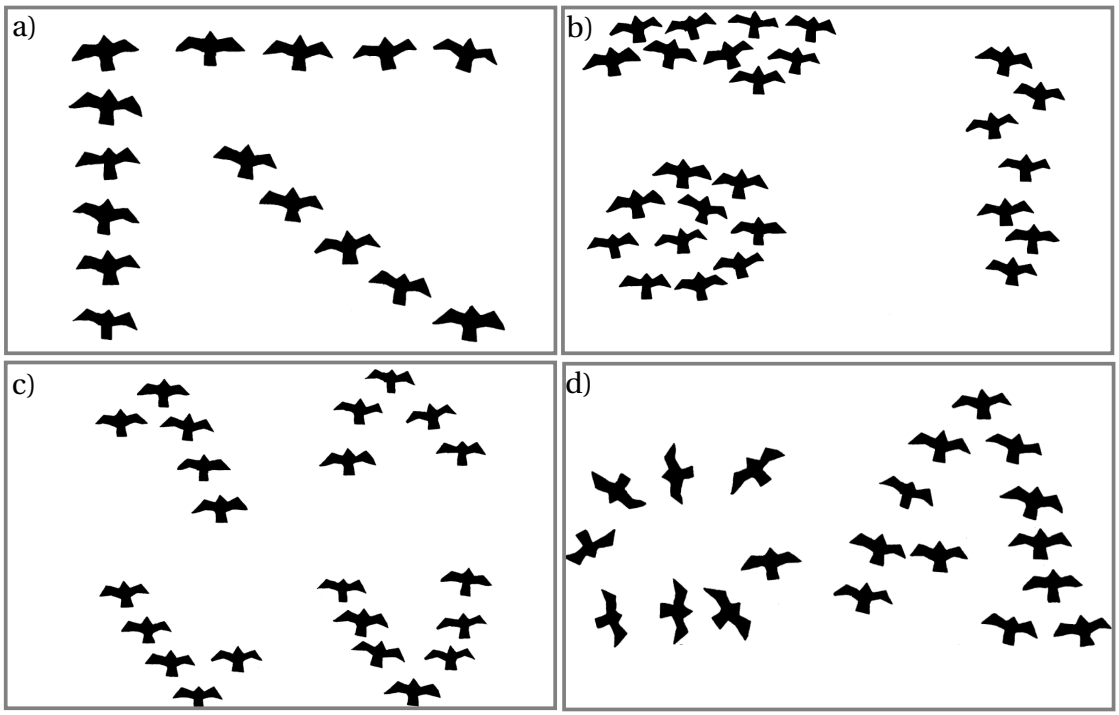

Figure 1.11.: Bird flock flight formations. a) Line formations, b) Cluster formations, c) Angled line formation. d) complex formations (reprinted from Heppner 1974).

Potts (1984) discussed collective movement of sandpiper flocks, initiated from observation of high-speed film. It was proposed that any individual of the flock could start a coordinated movement, which propagated through the flock as a ripple originated from that individual. This study mainly focused on the positional effects of an anti-predatory vigilance approach to the flock. Beauchamp (2003) evaluated this vigilante effects as a function of group size. It presented the effects of foraging speed, predatory detection, predatory protection and reaction speed. And one issue became apparent, the necessity of more empirical information.

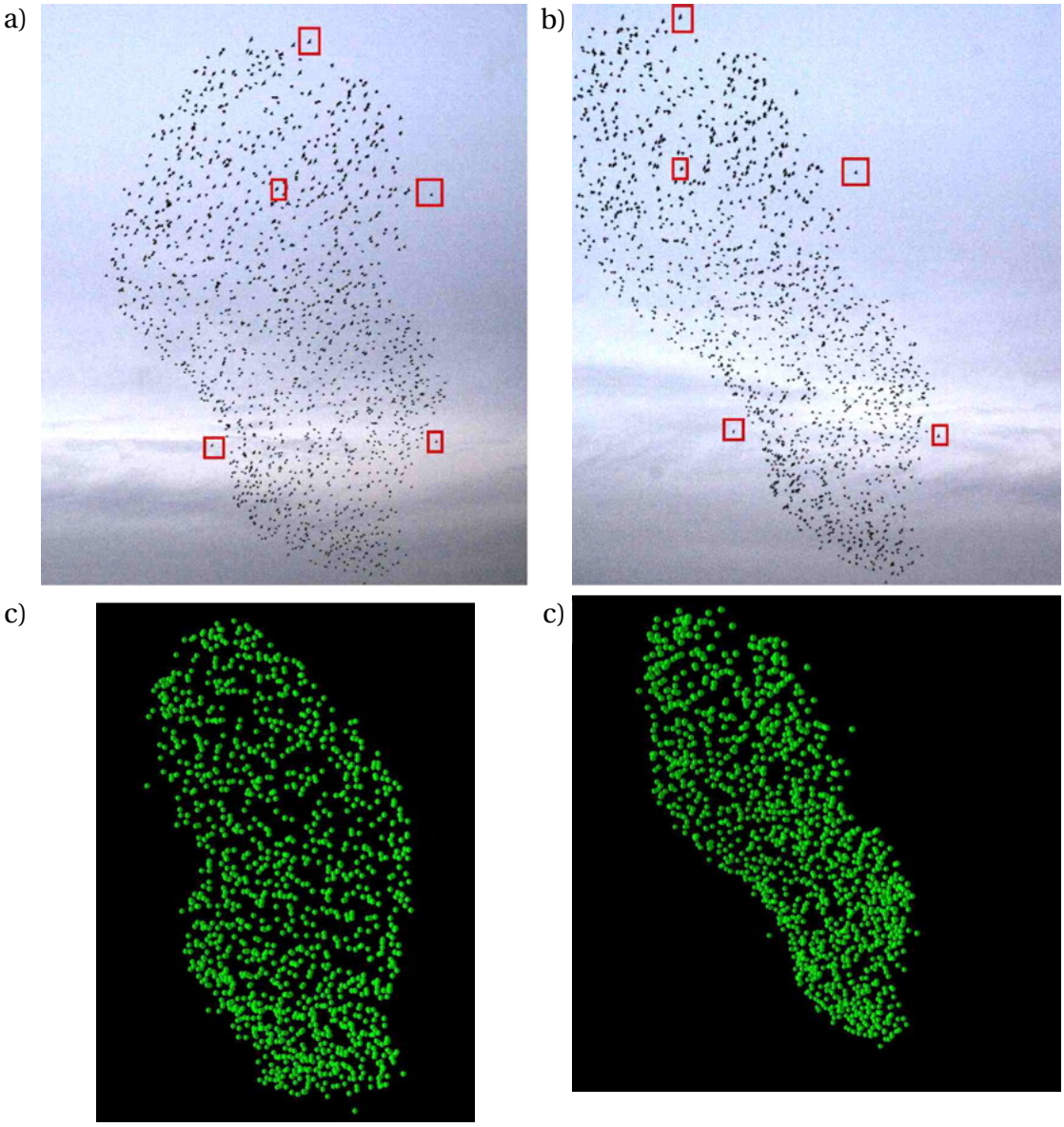

Figure 1.12.: Flock of starling recorded on the wild. Individual recogtinion between a) and b). Coordinate reproduction of the state of the flock on a) in c) and b) in d). Reprinted from Ballerini et al. 2008b.

The STARFLAG system (Cavagna et al. 2008a; Cavagna et al. 2008b) greatly contributed to a fast, cheap, and reliable data acquisition technique to obtain flight coordinates of thousands of flocking birds (Fig. 1.12). It presented a stereoscopic camera approach. This method allowed to study flocks from a topological point of view. Ballerini et al. (2008b) presented the reconstruction of the 3D data with an analysis on the range of interaction between individuals, showing that topological distance describes flocking behavior. Later, Cavagna et al. (2010) presented the response of the flock to an external perturbation. The study proposed that the collective response could be described by scale-free behavioral correlations. Based on this description, it presented a behavioral small correlation decay, that hinted to a limited interaction range between individuals.

1.2.5. Lessons of observing nature

The ubiquity of collective motion allows for analysis across several species. Papadopoulou et al. (2023b), presents a study of collective motion that analyses several data sets of four different species (sticklebacks, pigeons, goats and chacma baboons). They used a mix of local (nearest neighbor distance and bearing angle) and global (polarization and group shape) descriptors to compare the flocks when the animals presented collective motion (Fig. 1.13). This study presents the appearance of similar behavioral patterns across multiple species. From the multitude of examples of collective motion, the following remarks can be made about collective motion patterns:

- Neighbors motion and a tendency to imitate this motion, is the main reason for collective motion.
- Very similar behaviors occur in systems of very different origin. Opening the possibility for universal classes of collective motion patterns.
- Boundary conditions are a significant player in controlled collective motion reproduction.

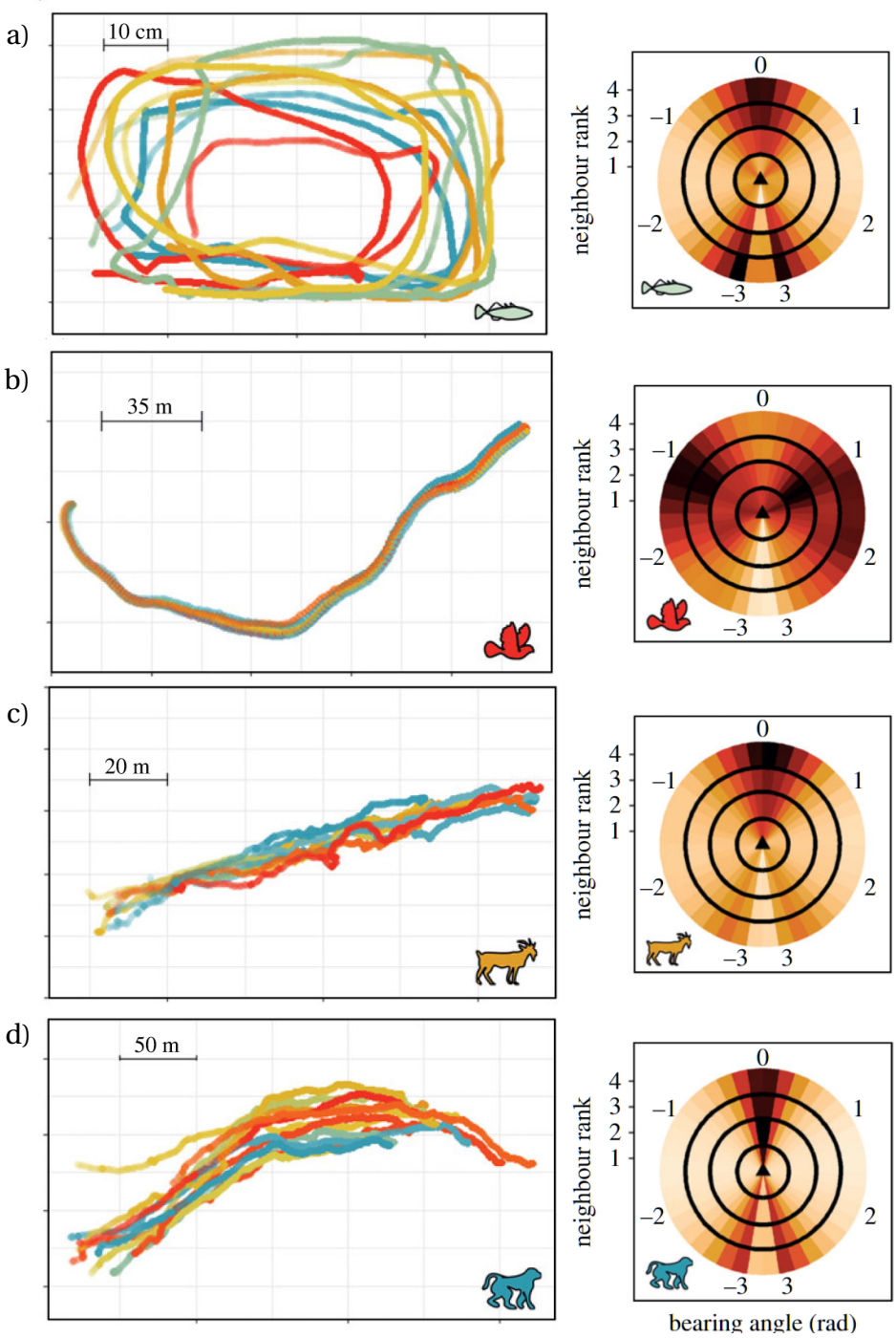

Figure 1.13.: Multiple species collective motion. a) Sticklebacks, b) Pigeons, c) Goats, d) Chacma baboons (reprinted from Papadopoulou et al. 2023b).

1.3. Modeling Nature

Section 1.2 already examined different species and examples of collective motion. Then the following questions become: Why collective motion occurs? How can we describe collective motion? Both questions still are a highly debatable subject, because there are many possible answers to it. There are biological (Gordon 2014), physiological (Germine 1997), behavioral (Russell et al. 2017), social (Bode et al. 2011b) even psychological (Sasaki et al. 2018) explanations for collective motion. But there is one description that is irrefutable, and is to explain the collective motion of a group as if it was a thermodynamic system. A statistical approach that aims to explain the phenomena from an already understood problem. To be more specific, it aims to describe the interactions inside the flock as if every individual would be a particle and explain the global order as a phase, similar of the phases of matter. Furthermore, this led to several models and numerical simulations of those models, that we will explore in the following sections.

1.3.1. Self-Propelled Particle Models

The first simulation of collective motion is attributed to C. W. Reynolds (1987). In this study, a simple model was proposed that allowed the natural visual appearance of coherently flying objects, referred as "boids"[2]. These "boids" were self-propelled particles (SPP) with a pyramidal geometry. The model formulated a set of three behavioral rules (Fig. 1.14)(taken from C. W. Reynolds 1987):

- "*Collision Avoidance*" explained as "*...the urge to steer away from an imminent impact. Static collision avoidance is based on the relative position of the flockmates and ignores their velocity.*"
- "*Velocity Matching*" explained as : "*...boid's awareness of another is based on the distance and direction of the offset vector between them... velocity matching is based only on velocity and ignores position.*"
- "*Flock Centering*" explained as "*...makes a boid want to be near the center of the flock. Because each boid has a localized perception of the world, "center of the flock" actually means the center of the nearby flockmates.*"

[2]This objects give the alternative name to this model (Boids model)

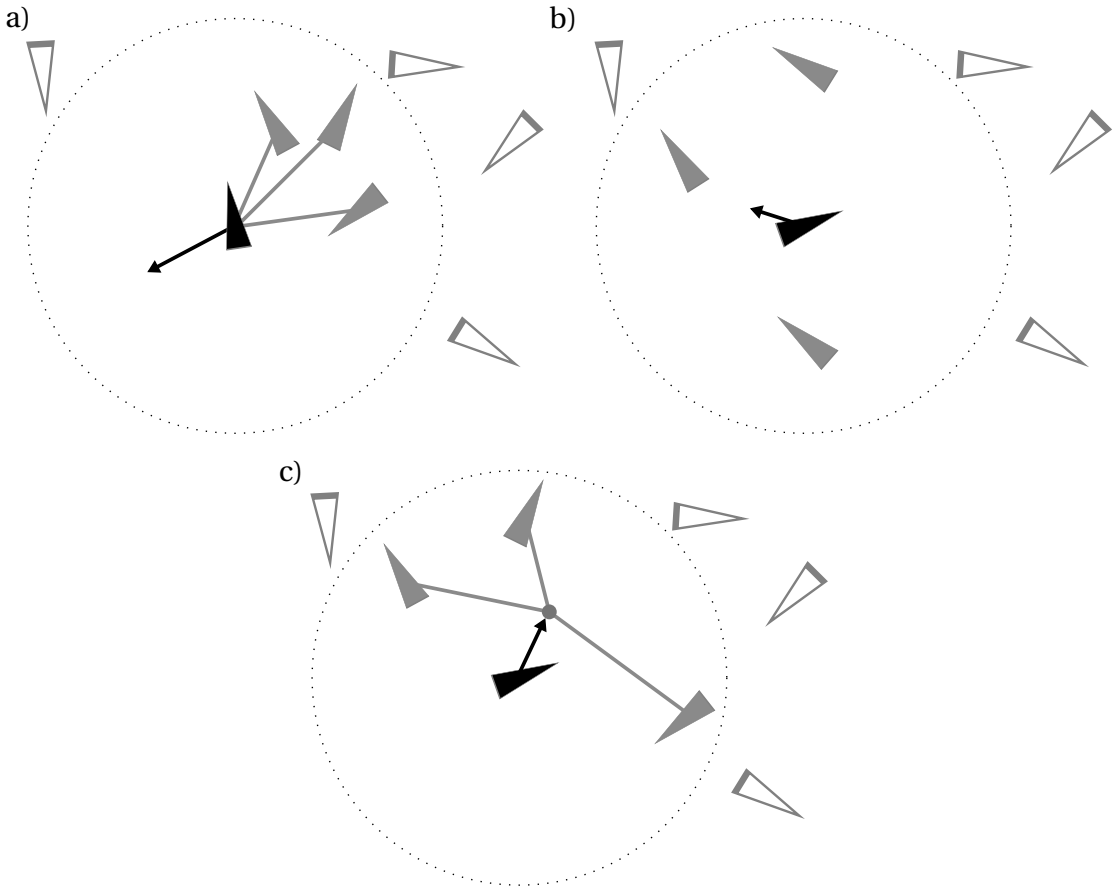

Figure 1.14.: *Boids* Model behavioral rules. a) Avoidance, b) Alignment, c) Attraction (reprinted from C. W. Reynolds 1987).

The collision avoidance and velocity matching present a complementary relationship, as a correctly implemented matching would maintain the distance between particles and the collision would only take action if there was a conflicting behavior that disrupted a steady state. The flock centering would act strongly to those on the edge of the group, as those in the center would be on an area of homogeneous density (their center would coincide with themselves). These are simple behavioral rules, where the specific of the interactions are left open to interpretation or to be parameterized.

Then a statistical physicists type analysis was introduced by Vicsek et al. (1995), widely known as the "*Vicsek's Model*". They proposed a model of particles bounded to a plane that averaged the velocities in an area around each individual (*zone of attraction* zoa_i). Each particle moves at a fixed absolute speed (U). Furthermore, the model was complemented by introducing perturbations (η), which modeled the many stochastic and deterministic elements that affect the motion of any given individual. These perturbations are assumed as an additional random angle to the average direction of the particle. The model is described by Eqs.1.1-1.2

$$\mathbf{x}_i(t+1) = \mathbf{x}_i(t) + U\mathbf{e}_i(t)\Delta t \tag{1.1}$$

$$\theta_i(t+1) = \overline{\theta(t)}^{zoa_i} + \eta \tag{1.2}$$

where $\overline{\theta(t)}^{zoa_i}$ denotes the average direction of all particles in the area zoa_i including the i-th particle, $\mathbf{e}(t) = [\sin\theta_i, \cos\theta_i]$ and η a random number with uniform distribution between $[-\sigma/2, \sigma, /2]$. With σ, being the standard deviation of the particles. The model proposed on Eqs.1.1-1.2, was subject to numerical simulations changing the values of σ and U of a flock size of 300 particles in an area of a square of side L with periodic boundary conditions. The speed value was observed to not affect the results in the range of $0.003 < U < 0.3$. Consequently, for all the rest of the numerical simulations, U was set to 0.03. By varying the area and noise level, they found that with a big square and low noise ($L = 25, \sigma = 0.1$) the particles tend to move in smaller groups in random directions (Fig. 1.15a), at smaller squares and high noise ($L = 7, \sigma = 2$) the particles move randomly (Fig. 1.15b) and with a small square and low noise ($L = 5, \sigma = 0.1$) all the particles move in the same direction (Fig. 1.15c).

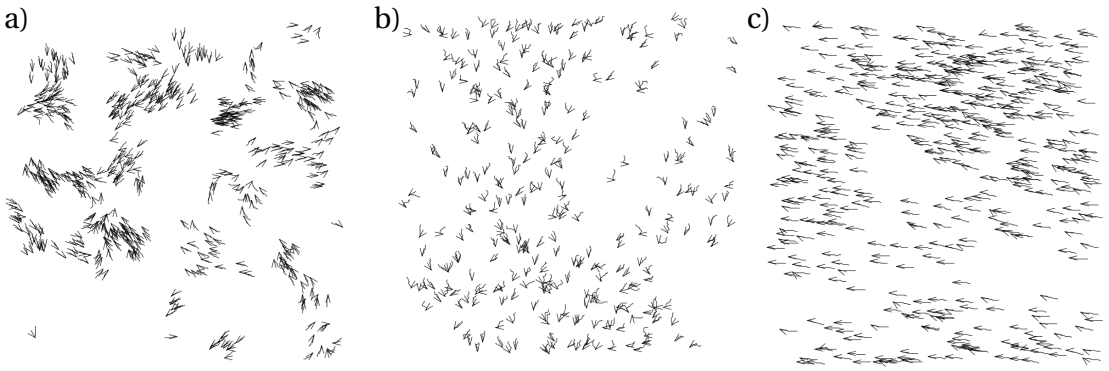

Figure 1.15.: *Vicsek* Collective motion reproduced. a) Organized clusters, b) non-ordered cohesion, c)flock-wide alignment (reprinted from Vicsek et al. 1995).

These two models (Boids and Vicsek's) are the basis for the statistical description of collective motion and opened the exploration of different implementations of SPP models. Shimoyama et al. (1996) evaluated the model without perturbations, Baglietto et al. (2008) evaluated a more complex region of interest and later Baglietto et al. (2008) evaluated the phase transitions of the Vicsek model. Kulinskii et al. (2009) explored a decaying velocity of the boids model. Mishra et al. (2010) explored the fluctuation on the phases reproduces by the Vicsek model. Grégoire et al. (2004), Levine et al. (2000), and Nagy et al. (2007), all presented different exploration of the parameter space of the models.

1.3.2. 3A models

Reynolds proposed a behavior to reproduce the effect of collective motion, Vicsek presented a simple parameterized mathematical model to order the flock as seen in nature. Couzin et al. (2002) mixed both, they proposed a model that used the three behavioral rules (Attraction, Avoidance, Alignment *3A*) of Reynolds, implemented with a parameterized mathematical model of areas of interest and a "blind spot", Fig. 1.16.

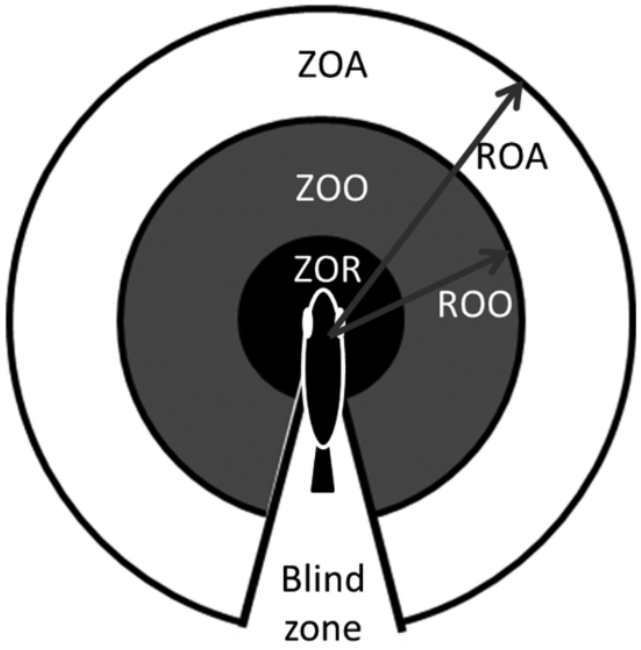

Figure 1.16.: Zones of interactions and blind zone. The Closest zone defined by a radius (*ROR*), enforces repulsion/avoidance (*zor*). The second zone defined by a radius (*ROO*), enforces orientation/alignment (*zoo*). The last zone defined by a radius (*ROA*), enforces attraction (*zoa*) (reprinted from Delcourt et al. 2016.

Couzin's model presents the behavioral rules as a set of three zones with distinct characteristics. The Closest zone, defined by a radius (*ROR*), enforces repulsion/avoidance (*zor*) for all particles as:

$$\theta(t)_r = -\sum_{j\neq i}^{n_r} \frac{\mathbf{r}_{ij}(t)}{\|\mathbf{r}_{ij}(t)\|} \tag{1.3}$$

$$\mathbf{r}_{ij}(t) = \frac{\mathbf{x}_j - \mathbf{x}_i}{\|\mathbf{x}_j - \mathbf{x}_i\|} \tag{1.4}$$

where $\mathbf{r}_{ij}(t)$ is the interdistance vector between the particle i and particle j. $\mathbf{x}_i$ is the absolute position of particle i.

Unless there are no other particles in the *zoa*, the other zones will be ignored. The avoidance zone priority ensures that the model is collision free. The second zone,

defined by a radius (*ROO*), enforces orientation/alignment (*zoo*) for all particles in the area (n_o) as:

$$\theta(t)_{\parallel} = \sum_{j=0}^{n_o} \frac{\mathbf{e}_j(t)}{\|\mathbf{e}_j(t)\|} \tag{1.5}$$

where $\mathbf{e}_j(t)$ is the heading unit vector of the particle i.

The area of orientation is a disk with minor radius of *ROR* and mayor radius *ROO*. The last zone, defined by a radius (*ROA*), enforces attraction (*zoa*) for all particles in the area (n_a) as:

$$\theta(t)_{\odot} = \sum_{j\neq i}^{n_a} \frac{\mathbf{r}_{ij}(t)}{\|\mathbf{r}_{ij}(t)\|} \tag{1.6}$$

The area of attraction is a disk with minor radius of *ROO* and mayor radius *ROA*. The update of the particles is defined as:

$$\theta(t,\tau) = \begin{cases} \theta(t)_r & n_r > 0 \\ \theta(t)_{\parallel} & n_r = 0 \,\&\, n_o > 0 \,\&\, n_a = 0 \\ \theta(t)_{\odot} & n_r = 0 \,\&\, n_o = 0 \,\&\, n_a > 0 \\ 0.5\theta(t)_{\parallel} + 0.5\theta(t)_{\odot} & n_r = 0 \,\&\, n_o > 0 \,\&\, n_a > 0 \end{cases} \tag{1.7}$$

Couzin's model presented a more complex interactions between the behavioral rules of Reynolds, the statistic of Viscek and expanded both. Having the best of both models allowed to have complex behaviors reproduced by this kind of model, which will be referred as "3A models"[3].

By changing the radius of effect of the zones, 3 phases were reproduced, swarm, milling (vortex) and school (Fig. 1.17b). These three phases can also be seen naturally occurring animal collective motion (see Section 1.2). Gautrais et al. (2008) performed another set of experiments in 2D where the same phases were reproduced (Fig. 1.17a).

Delcourt et al. (2016) they evaluated which phase could be reproduced by changing the size of *ZOO*. By doing so, the phase changes from the swarming when there is no alignment. Milling was reproduced when the area of alignment is small and finally school, when the alignment area is big. Furthermore, they explore the effects of changing the relationship presented on Eq. (1.7). This relationship induces a phase change between school and swarm. Reproducing the former when the relationship favors alignment, the latter when it favors attraction. If the relationship is maintained similar, the flock stays in a bistable phase. This bistability presents an equally probable school or swarm reproduction, alternating between both.

[3] 3A due to having Avoidance rule, Alignment rule and Attraction rule. 3A

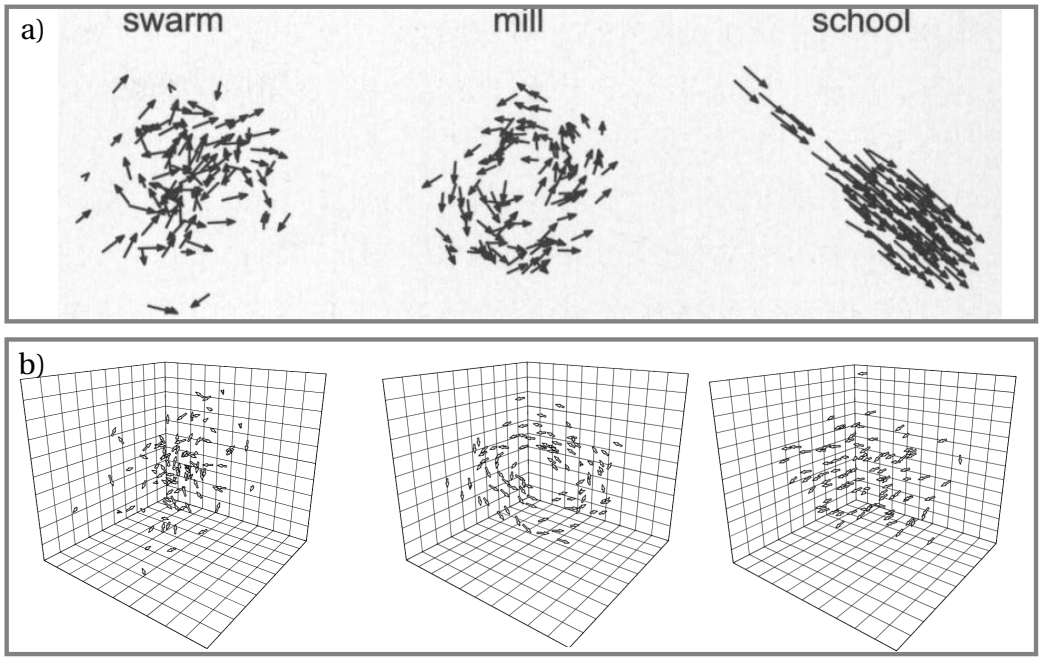

Figure 1.17.: Main collective motion phases: swarming, milling, schooling. a) 2D case (reprinted from Gautrais et al. 2008), b) 3D case (reprinted from Couzin et al. 2002).

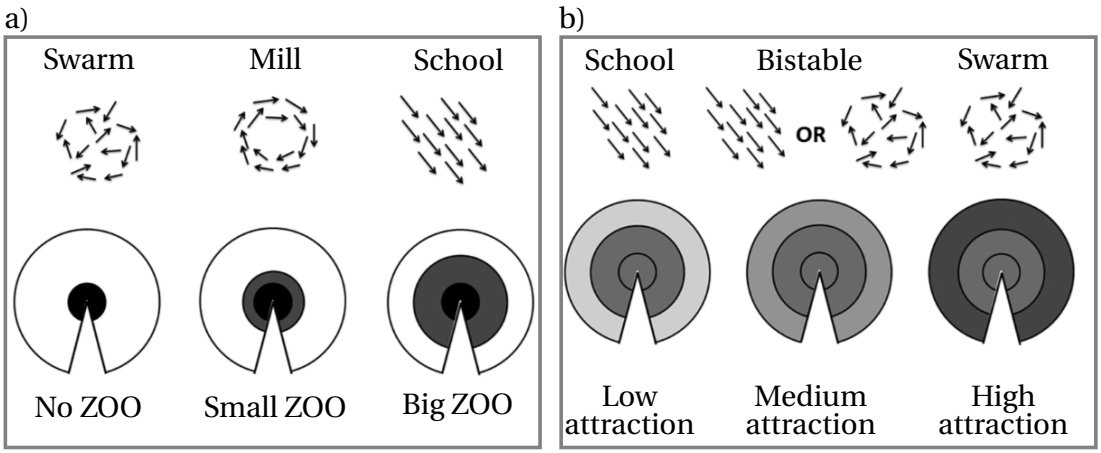

Figure 1.18.: Effect of zones of interactions on collective behavior reproduced. a) Effect of zone of orientation size, b) Effect of relationship between orientation and alignment (reprinted from Delcourt et al. 2016).

Gautrais et al. (2009) introduced a model to the perturbations and a component to model the interactions with the wall. It presented the non-collective movement as a turning walker described as a Wiener process and presented the wall as a repulsive force that interacted with the particle as a mirror would. The wall interaction allowed

the comparison between experimental data from natural flocks on controlled environments and simulation results. Then Gautrais et al. (2012), updated the interaction of each individual with the rest of the flock interaction. With this change the rules were no longer enforced based on regions of effect but to instead take into account just the closest neighbors (Voronoi neighbors) (Fig. 1.19).

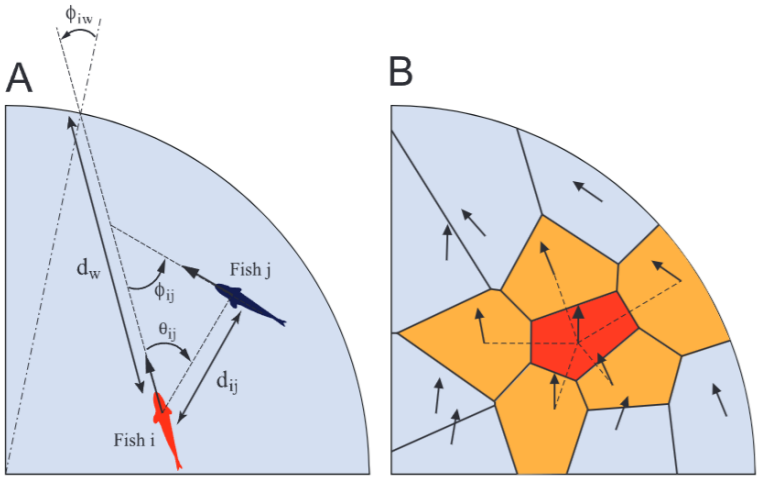

Figure 1.19.: Model based on neighbors. a) Notations used, b) Depiction of the Voroni neighbors for a particle (reprinted from Gautrais et al. 2012).

The change from area effect to closest neighbors, allowed the introduction of a complete parametrization of the model in terms of the rule's strength and a new continuous form of blind spot. The new blind spot function allowed to increase the effect of particles along the path of the individual while reducing it for those in the back. The model is updated using Eqs. (1.8)-(1.9).

$$\dot{\theta}_i^* = k_W \frac{sgn(\phi_{iW})}{\mathbf{d}_{iW}} + \frac{1}{N_i} \sum_{j \in V_i} [k_\parallel U \sin\phi_{i,j} + k_\odot \mathbf{d}_{ij} \sin\theta_{ij}] \tag{1.8}$$

$$d\omega(t) = -U \left[\frac{dt}{\xi} (\omega(t) - b\omega^*(t)) + k_\eta \eta \right] \tag{1.9}$$

where the W index indicate the wall interaction, θ the heading of a particle, ϕ_i, j the position of the particle j referred to the heading of particle i, $k_\parallel$ is the strength of the alignment rule, $k_\odot$ is the strength of the attraction rule, k_η is the strength of the noise and ξ is the time scale. The blindspot b is defined as:

$$b_\epsilon = 1 + \cos\theta_{ij} \tag{1.10}$$

Gautrais et al. (2012) took the proposed model on Eqs. (1.8)-(1.9) and evaluated the attraction and alignment parameter space $(k_\parallel, k_\odot)$ based on two global metric parameters. These are: P the polarization of the flock or how aligned the particles are (Eq. (1.11)), and M the milling of the flock or how much rotation there is in the

flock (Eq. (1.12)). Moreover, an evaluation on how these metrics associated with the parameter space and phase reproduction. The model reproduced swarming, milling, schooling and, it was found that in between the areas of swarming and milling there was a transition area of a bistability state (Fig. 1.20)

$$P = \frac{1}{N}\left|\sum_{i=1}^{N} \mathbf{e}_i\right| \tag{1.11}$$

$$M = \frac{1}{N}\left|\sum_{i=1}^{N} \frac{\mathbf{r}_i \times \mathbf{e}_i}{\|\mathbf{r_i}\|}\right| \tag{1.12}$$

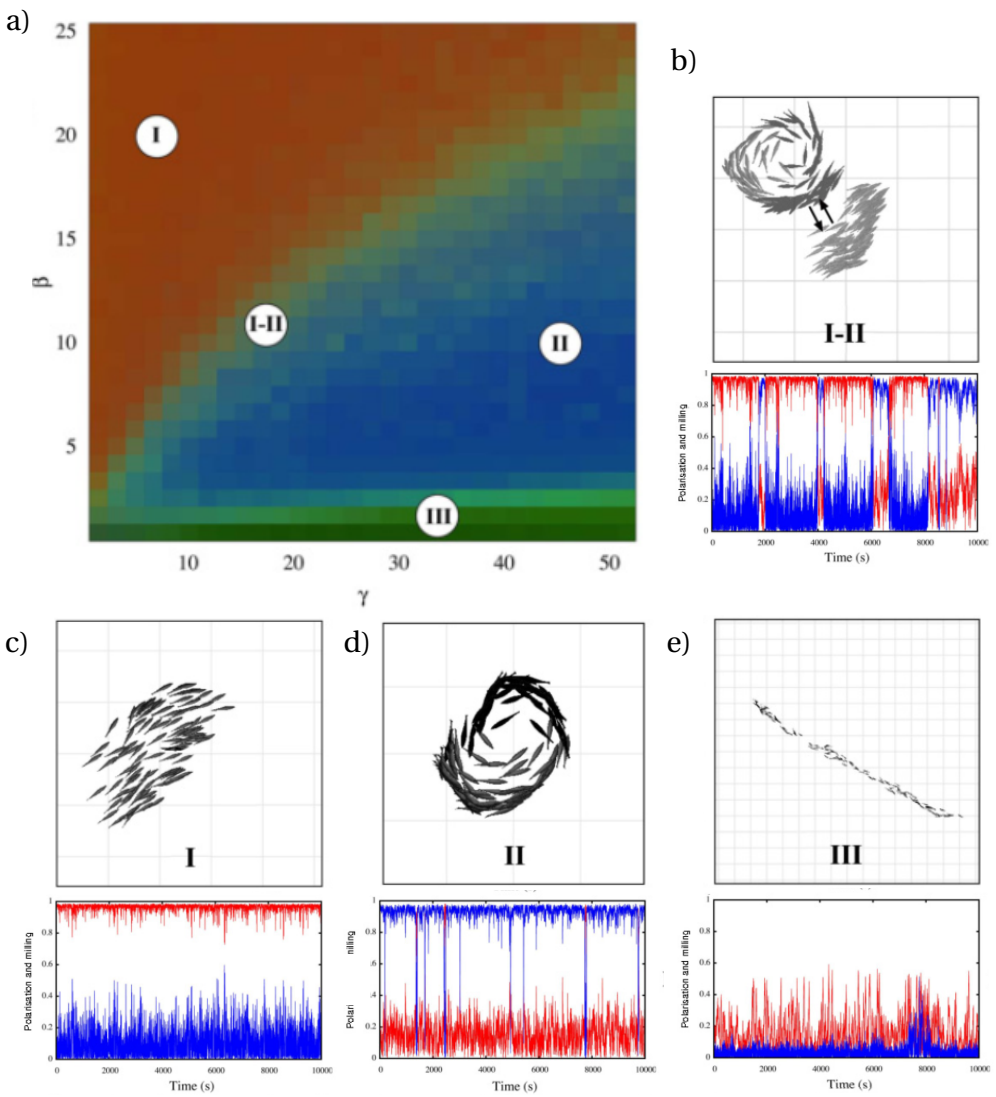

Figure 1.20.: Collective motion phases. a) Phase diagram, b) Bistable, c) Schooling, d) Milling, e) Swarming (reprinted from Calovi et al. 2014).

1.3.3. Non-aligning models

All the previous models had a component that tried to align all the individuals in the flock. It was a component that performed an average of the surrounding particles to decide where to align its own heading. Strömbom (2011) presented a model that did not contemplate any alignment, only the rule of attraction presented on Eq. (1.3) and the random process to integrate perturbations. Without any alignment, this model reproduced the same patterns seen on the other aligning models (Fig. 1.21). They also introduced a blind angle, which altered the patter from undirected mills to become directed, and a new "rotating chains" appeared (Fig. 1.21b). Without an explicit alignment rule, the presented results hinted at the role of the blind spot on aligning the particles on a local neighborhood.

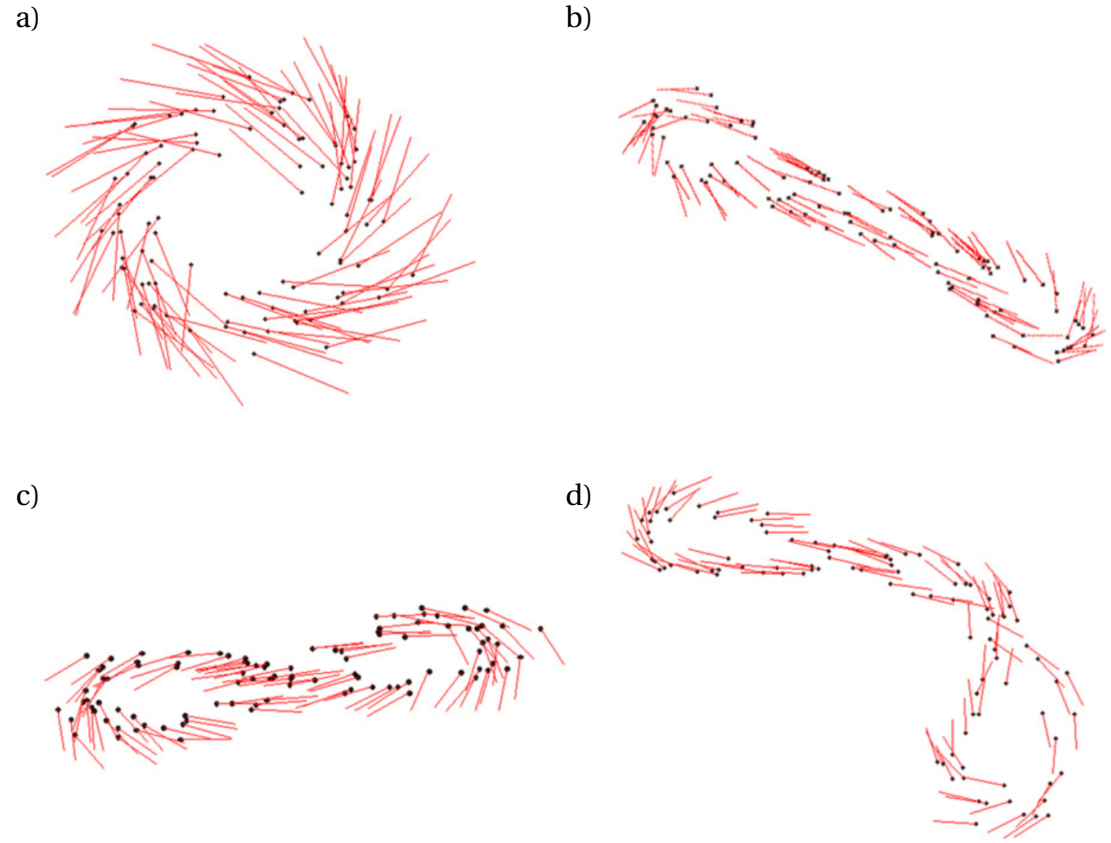

Figure 1.21.: Collective motion phases. a) milling, b) elongated milling, c) crossing milling. d) organized swarming (reprinted from Strömbom 2011).

Grossman et al. (2008) proposed a model without relying on alignment, instead assuming a specific form of inelastic collisions between particles. The model used self-propelled isotropic particles that moved and collided on a two-dimensional on a frictionless surface. Moreover, reflective boundary conditions were imposed, which led to the emergence of various collective behaviors (Fig. 1.22). Another analysis was to

change the boundary shape, taking circular, elliptical and squared shapes. With these different shapes, it was evaluated the effect on the collective behavior reproduced. The emergence of coherent collective motion in such a system is attributed to the fast alignments that inelastic collisions between isotropic particle provides. Note that these collisions do not conserve momentum, but result in a slight increase in velocity after each collision.

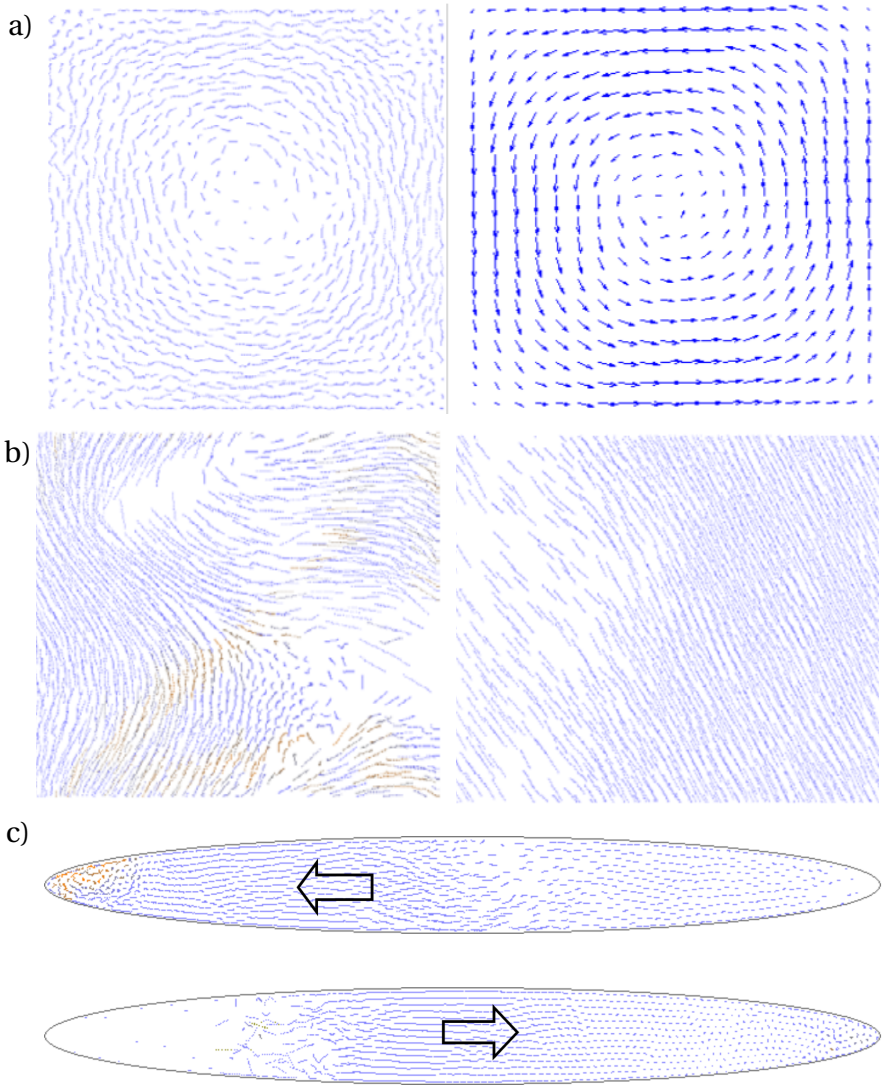

Figure 1.22.: Collective motion phases. a) milling, b) local schooling. c) wave-like movement (reprinted from Grossman et al. 2008).

1.3.4. Graph models

The introduction of a limited amount of interactions allowed the representation of the flock with a different set of descriptors. One fascinating way to describe it, is as a set of interconnected nodes. Each link between nodes (particles) would have an associated

heuristic, usually representative of the relative distance or velocities (Albert et al. 2002). The node presentation generate dynamic networks, which are typically referred to as a "switching topology" network. Jadbabaie et al. (2003) used a flock described using this kind of network to implement an attraction rule. Using a switching topology network, they were are able to reproduce similar emerging patterns as those showcased by traditional SPP models. Shang et al. (2014) presented a study to find the effects of these topological interactions on ordered movement in collective motion. The inter-flock interactions were modeled as graphs or networks instead of the traditional areas seen on SPP models (Fig. 1.23). The focus of this study was on reaching an ordered state -finding consensus-within the flock as a function of the number of neighbors each individual interacted with. Regardless of the amount of neighbors you consider, a theoretical consensus is mostly possible[4]. The amount only affects the time to reach consensus. Both theoretical hypothesis were confirmed by analytical and numerical analysis, finding that the rate of convergence approached the optimal point with 10 interacting neighbors.

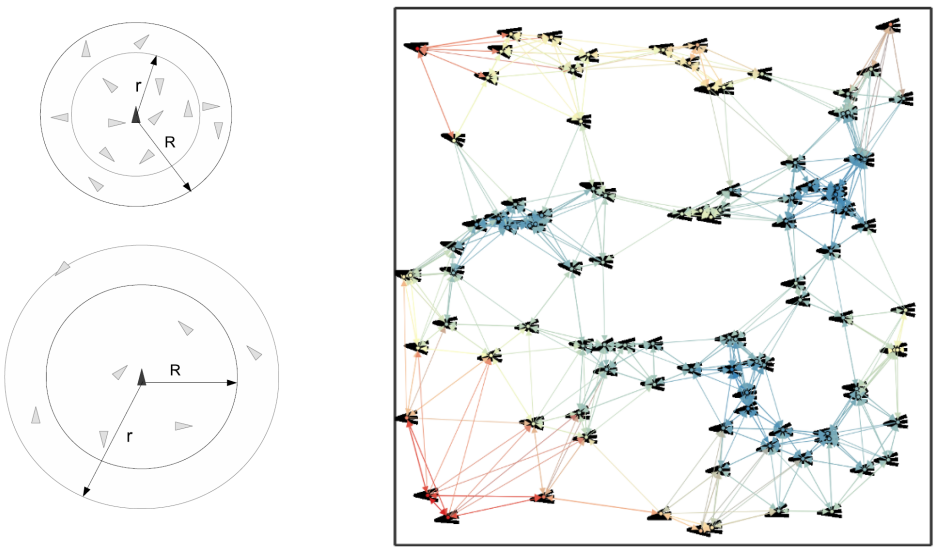

Figure 1.23.: left traditional approach, right graph network approach (reprinted from Shang et al. 2014).

The information diffusion of a graph representation, mimics the information distribution on a flock, with the added benefit of ignoring metric dependent measurements and leaning towards topological information. These kinds of graph-like formulations are mostly seen with implementation of collective motion with robotic flocks. The next section will explore robotic collective motion.

[4]This is true if the graph is connected (for undirected graphs) or as an spanning tree (for directed graphs)

1.4. Robotic collective motion



Figure 1.24.: Some robotic platforms for collective motion. (a) Jasmine Kornienko et al. (2005), (b) Alice Caprari et al. (2005), (c) Kilobots Rubenstein et al. (2014), (d) E-pucks Mondada et al. (2009), (e) Swarm-bots Mondada et al. (2005) and (f) Swarmanoid Dorigo et al. (2013) (reprinted from Dorigo et al. 2021).

Autonomous robots that act to reach a desired outcome without human intervention has been the aim of researchers. And having not just one robot act, but a flock towards a goal has an even broader appeal (Fig. 1.24). The research into robotic collective motion can be traced more than 30 years, Kernbach (2013) presents a great summary of the topic that expands on the limited scope of this section. On this section some examples of collective motion on robotic swarm research will be showcased.

One of the earliest studies towards collective group of robots can be seen in Mataric (1994). In this study simple group behaviors (Wandering, Follow, Disperse, Aggregate, Homing) were combined to move collectively, to forage for resources or to return to a designated rest spots (Fig. 1.25).

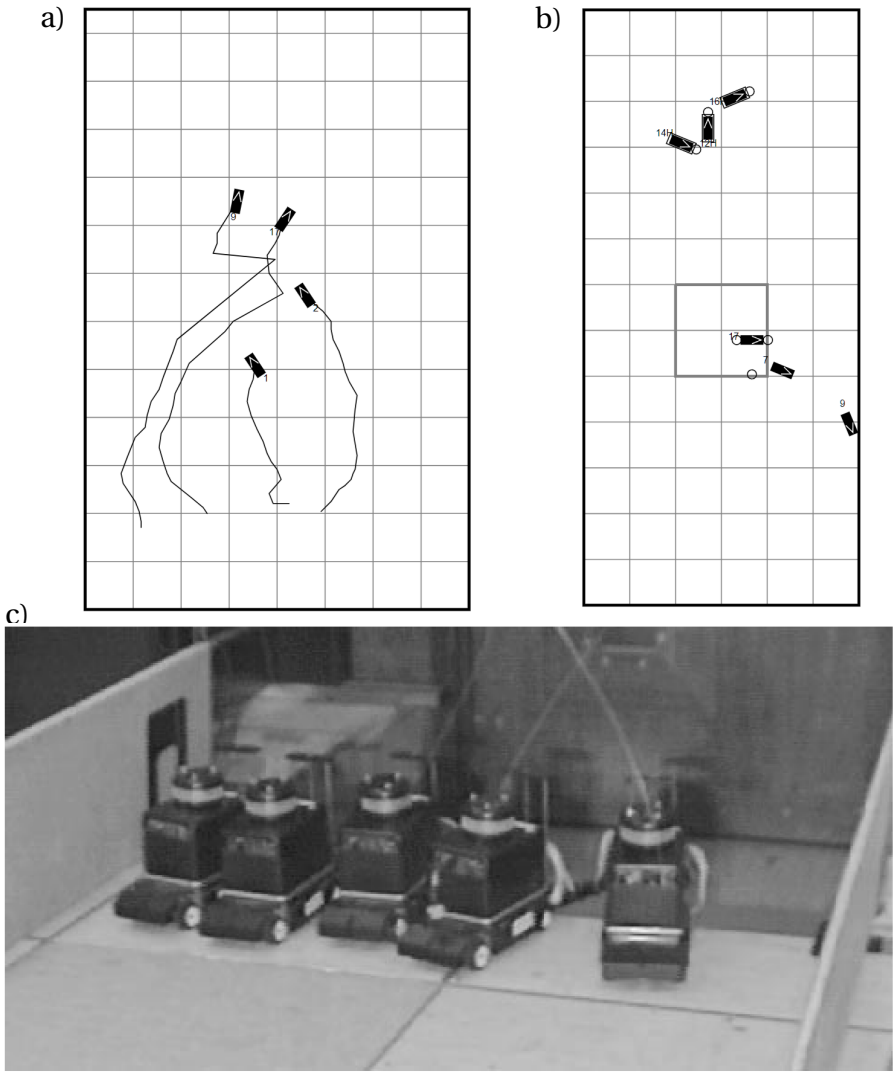

Figure 1.25.: Behavioral collective motion on robots. a) Coordinated movement, b) Foraging, c) Docking (reprinted from Mataric 1994).

Hayes et al. (2002) proposed a collective model with two of the three behavioral

rules of Reynolds: avoidance and alignment. The proposed algorithm, assumed that the robots were able to sense the distance and direction of their neighbors within a predefined range. Then each robot would compute the flock center-of-mass based on the relative location of its neighbors. This approach was simulated, optimized and then verified on a 10-robot group. The robotic experimental verification had to calculate the inter-distance and direction of the robots using an overhead camera system and broadcast them to the robots. Holland et al. (2005) presented a similarly Reynolds-inspired model for unmanned aerial vehicles (UAV). They presented an approach that assumed each UAV had a wireless communication module and a ground-facing camera on-board. Then the UAV would capture an image of the ground and transmitted it to a host, which calculated and, sequentially, broadcast the inter-distance, direction and velocity of the UAV and its neighbors. The proposed broadcasting method allows the reproduction of swarming with this kind of drones. More recently, Verdoucq et al. (2023), implemented a full 3A model, that included a model of the wall and the effects of altitude within a cylindrical model. This study assumed that the drones would have information about the location, direction, and velocities of all its influential neighbors. With this minimal interacting model, the drones were able to reproduce schooling, swarming and milling (Fig. 1.26).

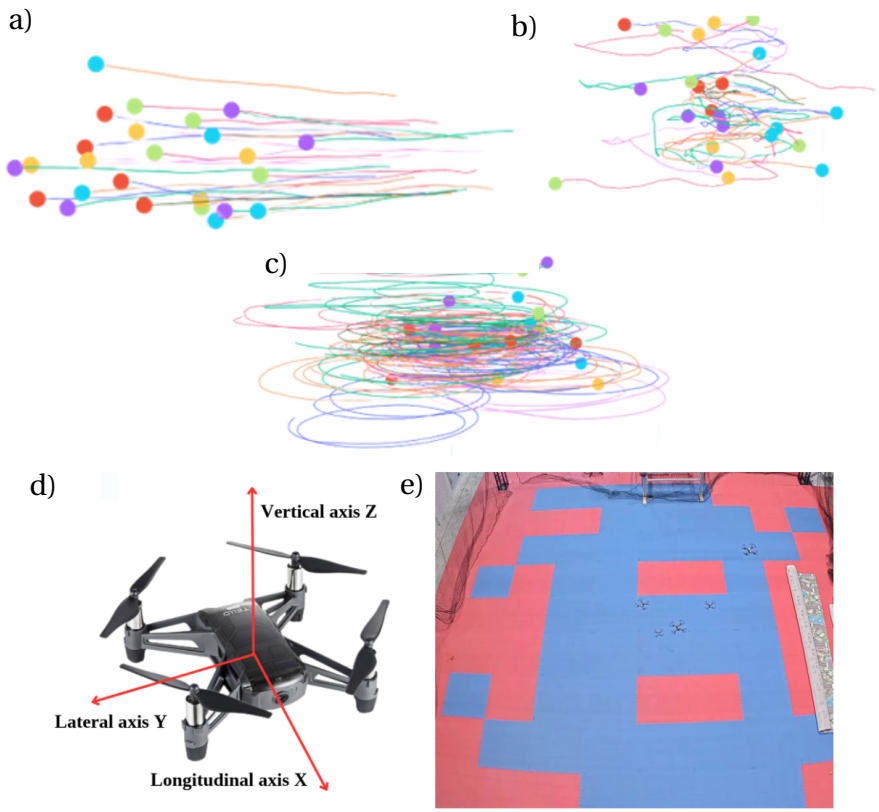

Figure 1.26.: 3A model in robots. a) Schooling phase reproduction, b) Swarming phase reproduction, c) Milling phase reproduction, d) Drone used, e) arena used (reprinted from Verdoucq et al. 2023).

Mateo et al. (2019) explored the influence of the network topology on collective motion (Fig 1.27). This study considered a model of distributed decision-making and evaluated the flock capability to follow a signal. The ability to follow a signal was evaluated for several varying network topologies and flock sizes. The study found that the swarm of robots topology affects the frequency at which the signal prompted a response. The emergent collective response to low frequency signals increases with the degree of the interaction network, while the opposite is true for the response to fast-changing ones. In other words, the more connected the flock is, the less agile it becomes.

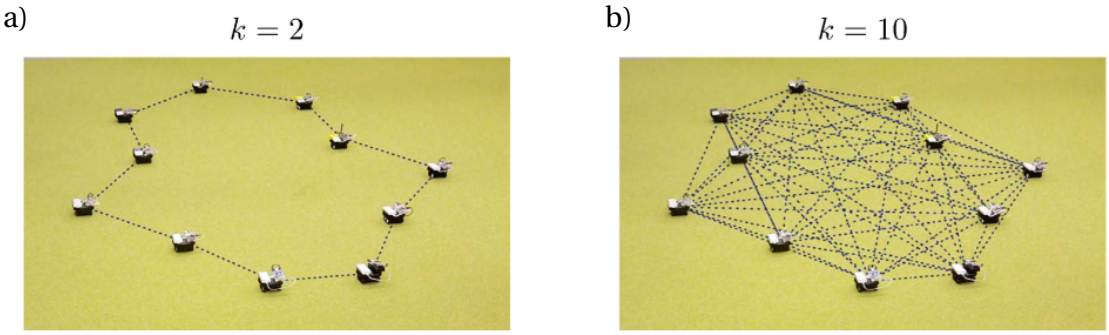

Figure 1.27.: Graph networks and robots. a) low interconnection between individuals, b) full interconnection between individuals (reprinted from Mateo et al. 2019).

Despite the years of research on collective motion, there is a miss match between natural occurring collective motion, collective motion models and robotic flocks. One of the possibilities to reduce this miss match is to use vision. Vision is widely used in nature, has been widely used on robots and has been the bases of several bio-mimetic and bio-inspired advancements. The following section will focus on the importance of vision, and visual information, on animals.

1.5. Vision and animals

The main animal groups that naturally display collective behavior are: insects -mainly known by swarming[5]-, fish -classic display of milling- and birds -well known schooling-. Then a brief look at the effect of vision on these animals groups must be done. The eyes are a part of the brain, and the eye itself is partially composed of neurons. Meaning that the eye is capable to both transmit stimulus to the brain and to perform low-level visual processing. On vertebrates, the light that enters the eye, is mostly diffracted by the cornea while it's focused by the lens to reach the retina, where it is detected by the rods and cones (specialized cells sensible to light). Rods are mostly

[5]Note, this "swarming" phase is a different concept to swarm robotics. Here, the swarming phase is an emerging behavior that occurs when the group moves in a non-organized motion.

active at low lights, while cones at high. Humans have three types of cones, that respond to different frequencies of the visible light spectrum, allowing for color vision.

Unlike humans, most insects have a completely different eye, the compound eye (the most common eye type in the animal kingdom according to Cronin et al. (2014)). This eye has evolved to such a level that a fly dedicates about two-thirds of their neurons to the processing of visual information (Strausfeld et al. 1976). A compound eye is composed of multiple *ommatidia*, knowing that an *ommatidium* is a crystalline cone, a, typically, hexagonal lens, and a *rhabdom*. A *rhabdom* is separated into 7 or 8 *rhabdomeres* that are surrounded by centerly migrated pigment cells (Kirschfeld et al. 1969). To characterize these photoreceptors, you can use two main optical parameters: the interommatidial angle ($\Delta\phi$), and the acceptance angle ($\Delta\rho$) (Fig. 1.28)

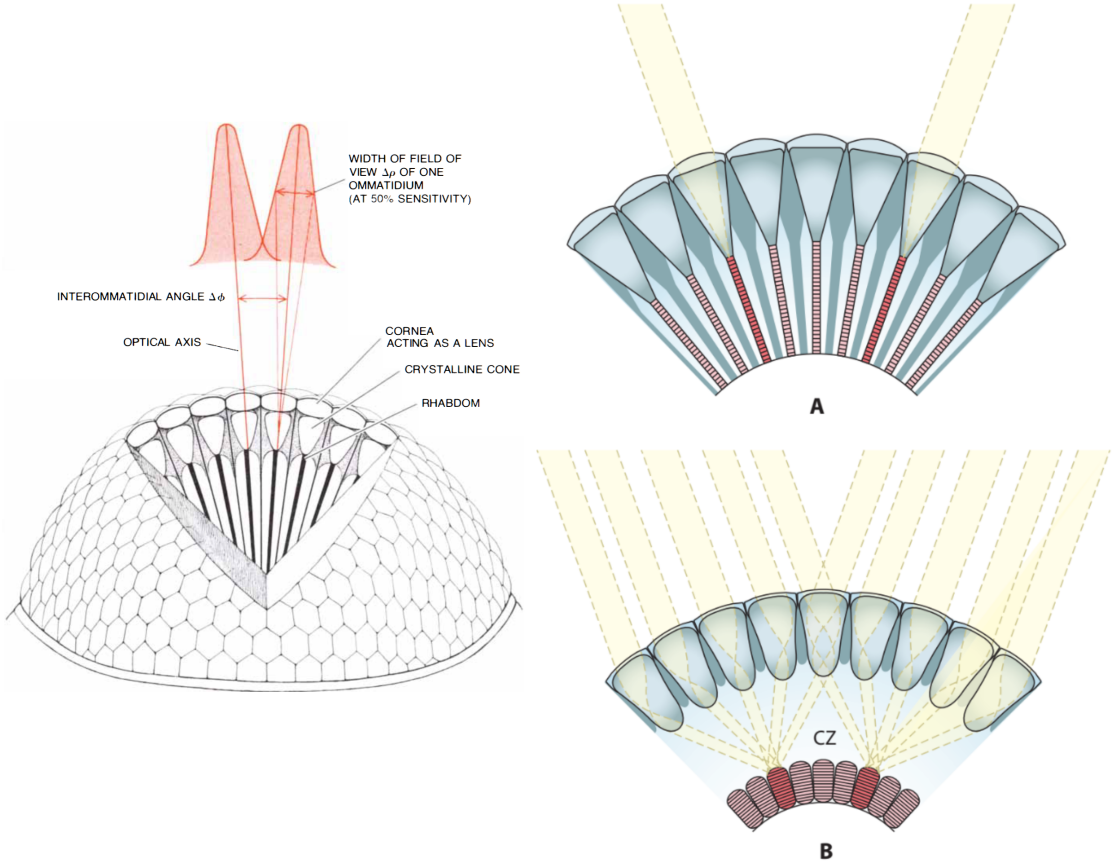

Figure 1.28.: Compound eye of insects. Left the general structure, right the types of compound eye (reprinted from Cronin et al. 2014; Horridge 1977).

There are two types of compound eyes. The *appositition*, where light that enters through a lens is guided to a unique *rhabdom*. And the *superposition* eyes, where the *ommatidium* are not separated with pigment cells, allowing the presence of a clear zone between the crystalline cones and the *rhabdoms*. This enables light from different lenses but same angular direction to excite a single *rhabdom*.

But there is no need to have specialized eyes to have incredible visual prowess. Birds, particularly predatory birds such as hawks and eagles, are highly renowned for their visual capabilities, capable of identifying small prey from long distances. Non-predatory birds (particularly passeriformes[6]) have laterally placed eyes and a single centrally located fovea[7] projecting into the lateral visual field (Fig. 1.29). With this kind of visual distribution, some species are able to detect predators and track their position and speed with high accuracy (Devereux et al. 2006). The detection is done using their lateral rather than their binocular visual fields. In other words, having a wider visual field is more relevant than binocular vision to avoid predations. An example of this effect can be seen on starlings (known for their collective motion) that have their fovea further forward, allowing a broader lateral visual field.

[6] perching birds, with four toes, three directed forward and one backward
[7] pitted invagination of the retinal tissue

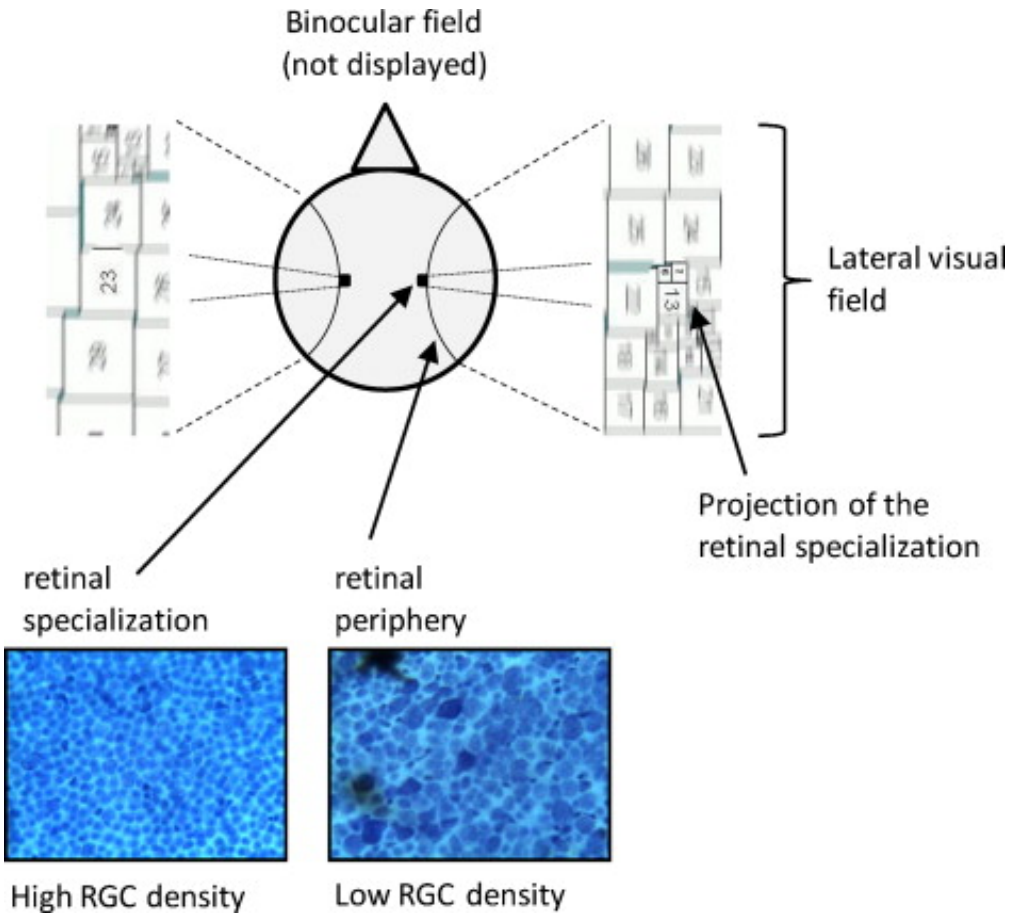

Figure 1.29.: Schematic representation of the projection of a central retinal specialization into the two lateral visual fields at a given head and eye position. Visual resolution is the highest at the retinal specialization, but decreases towards the retinal periphery seen as blurred vision due to lower density of photoreceptors and retinal ganglion cells (RGC) (reprinted from Fernández-Juricic 2012).

The broad field of view decreased the importance of ability to discriminate local changes in luminosity, known as contrast sensitivity. Ghim et al. (2006) presented a study that explicitly looks at the contrast sensitivity of starlings, among other bird species. The study suggested that these birds field of view are likely to group as a single optic object all birds have similar luminescence, even though shape recognition may allow them to be recognized as several birds.

Fish, have a similar eye to humans, with most fish species having a similar wide field of view of birds (Fernald 1988). The key difference, in the physiology of the eye, has to do with the frequency response of fish compared to non-aquatic animals. The light travels differently in water compared to air, this impacts the frequency at which each of the light sensitive cells interact (Douglas et al. 2012). As the light sensitivity of the fish eye is different, it also impacts the behavior of the fish. Lafoux et al. (2023) explores how different light intensities affect the collective behavior of fish, particularly the

species *Hemigrammus rhodostomus*. Using an experimental setup with controlled light variations, the study analyzes how the fish form and maintain groups under varying illumination levels. At low light levels, fish fail to establish cohesive groups, moving independently with minimal alignment or collective motion. As the light intensity increases, the fish gradually transition to more organized behaviors (Fig. 1.30a). Three distinct phases of group behavior were observed as the light intensity increased. Initially, the fish show disordered motion, followed by a phase of strong alignment at moderate light levels, and finally a stable, rotational "milling" behavior at higher illumination levels (Fig. 1.30b).

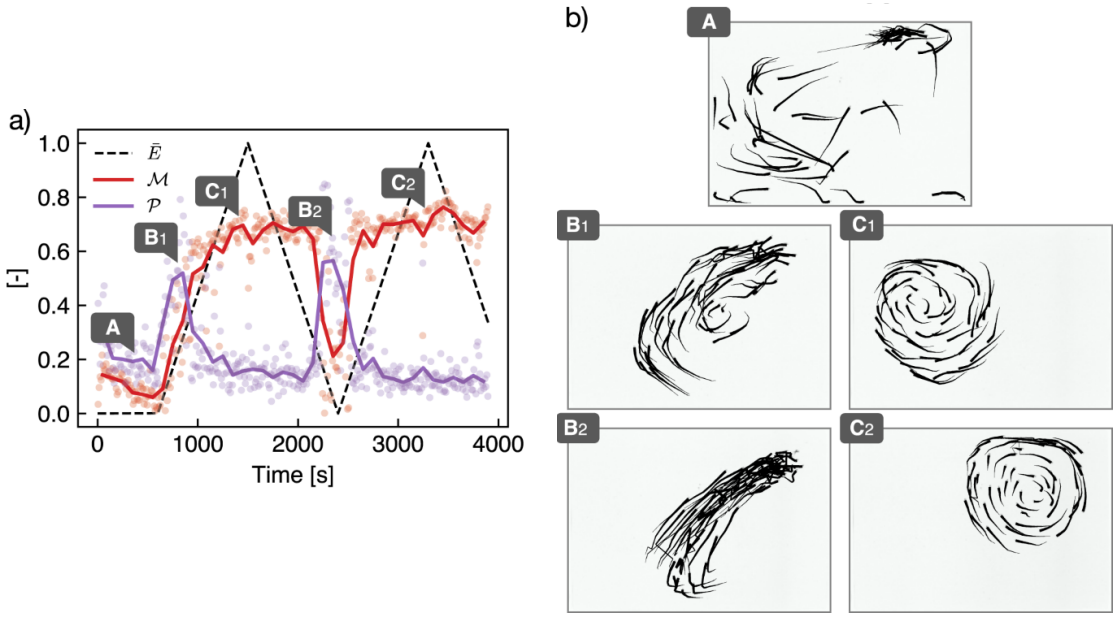

Figure 1.30.: Fish illumination collective response. a) Time series of the evolution of the collective metrics of milling (M) and polarization (P) correlated with the illumination (E), b) Observed phases (reprinted from Lafoux et al. 2023).

1.5.1. Optic flow

Optic flow is the perception of a retinal projection of apparent motion generated by relative motion between an observer and the environment (Gibson 1950). In short, is the perception of apparent motion generated by moving optic features on the retinal plane. Note that the changing optic features could be generated either by a moving environment, self-movement or both (Fig. 1.31). The apparent movement of optic features becomes an optical vector field orthogonal to the angular apparent movement vector at the feature location.

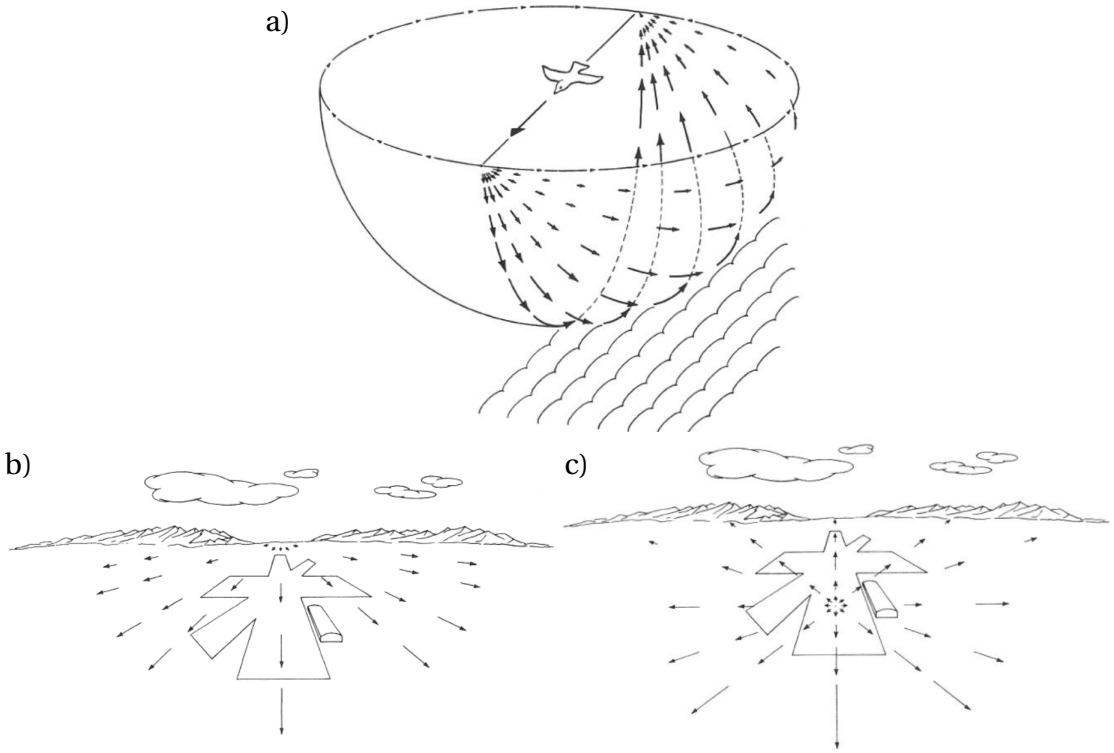

Figure 1.31.: Optic Flow examples. a) Optic flow generated by an observer movement, b) Schematic representation Optic flow generated by an observer movement parallel to the ground, c) Schematic representation Optic flow generated by an observer movement parallel heading to the ground (reprinted from Gibson 1950).

In this vector field exists spots where the magnitude of the vector becomes 0. If these spots align with an orthogonal direction of the apparent movement of an optic feature, they are called focus of expansion (FOE). The FOE can be divergent, when apparent movement is anti-parallel with the observer movement (as seen on the max elevation of Fig. 1.32b); or convergent apparent movement is parallel with the observer movement (as seen on the min elevation of Fig. 1.32b). Note that a pure rotation on an axis might yield a spot of 0 magnitude (Fig. 1.32c), but at the focal direction, denoted as F, there are not enough optical features to generate an optic flow. If we had infinite optical resolution, there would not be a 0 magnitude vector, unlike on a FOE.

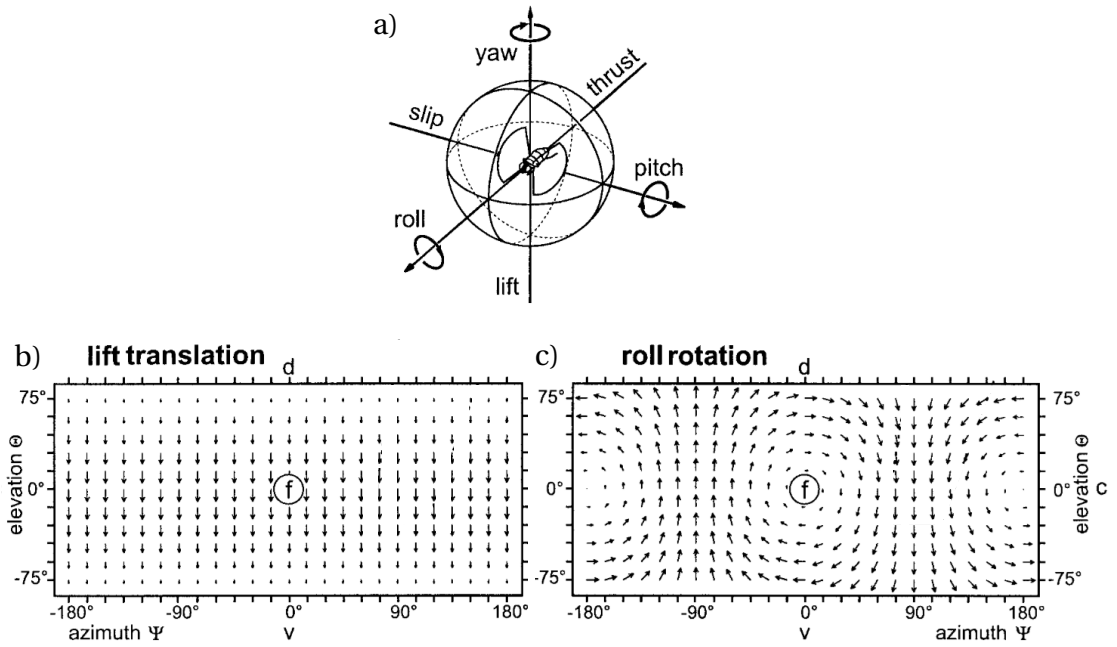

Figure 1.32.: Mercator projections of the optic flow vector field. a) Motions of a fly described by translation (thrust, slip, lift) and rotation (roll, pitch, yaw) around the 3 body axes, b) Optic flow vector field of a solely translational lift movement, c) Optic flow vector field of a solely rotational roll movement (reprinted from Krapp et al. 1998).

Koenderink et al. (1987) described the motion field or parallax $\mathscr{P}$ generated by a fixed "fiducial point" P (feature) with respect to a moving "vantage point" O (Observer), corresponding to the nodal point of the retina, as:

$$\mathscr{P} = \frac{1}{\|d\|}((T \cdot \hat{d})\hat{d} - T) - R \times \hat{p} \tag{1.13}$$

where d is the distance vector of the location of the feature P with respect to O. $\hat{d} = \frac{d}{\|d\|}$ is the unit vector representing the direction of d. And T and R are the translational and rotational vectors corresponding to the motion of the vantage point with respect to an inertial frame, respectively. Eq. (1.13), shows that the parallax motion can be decomposed by the linear and the rotational movement of the feature and observer. The linear, or translational, movement is inversely dependent on the distance from the observer to the feature, while the rotational is independent.

Although the parallax is a 3-D vector field, it can be described as an orthogonal 2-D vector field to $\hat{d}$. This orthogonal vector field is the optic flow. This allowed, the optic flow, to be traditionally represented using a Mercator projection, defined by an azimuth (Φ) and an elevation (Θ) components. The definition of the components allow calculating the optic flow of a single translational (Eq. (1.14)) or singular rotational (Eq. (1.15)) movement.

$$\mathcal{O}_r = \omega\sqrt{1-\cos^2\Psi\cdot\cos^2\Theta} \tag{1.14}$$

where ω is the angular velocity of the movement that generated the optic flow.

$$\mathcal{O}_\phi = \frac{U\mathbf{e}}{d}\sin\phi \tag{1.15}$$

where U is the magnitude of the linear velocity, $\mathbf{e}$ the unit vector of the linear velocity, d the distance between the feature and the observer and ϕ the retinal angle of the feature refereed to the heading of the movement.

The next challenge is to calculate the optic flow from available sensors, more specifically from digital cameras. This proposed another challenge, now the available optical features are bound to the geometry of the pinhole camera and to a finite temporal and spatial resolution. The pinhole camera presents a limited visual field that would make the recognition of a feature in different time steps a challenge. For example, Fig. 1.33 a) presents a line, with the perceived movement of this feature in the pinhole A being $\vec{P}$. Yet, the true movement of the feature is $\vec{R}$. This phenomenon is natural, as the pinhole limits information of the environment. And the issue extends when you have multiple features (Fig. 1.33 b). You can have a recognition of features P1 and P2 as Q1 and Q2 or Q2 and Q1, and consequently for even more features.

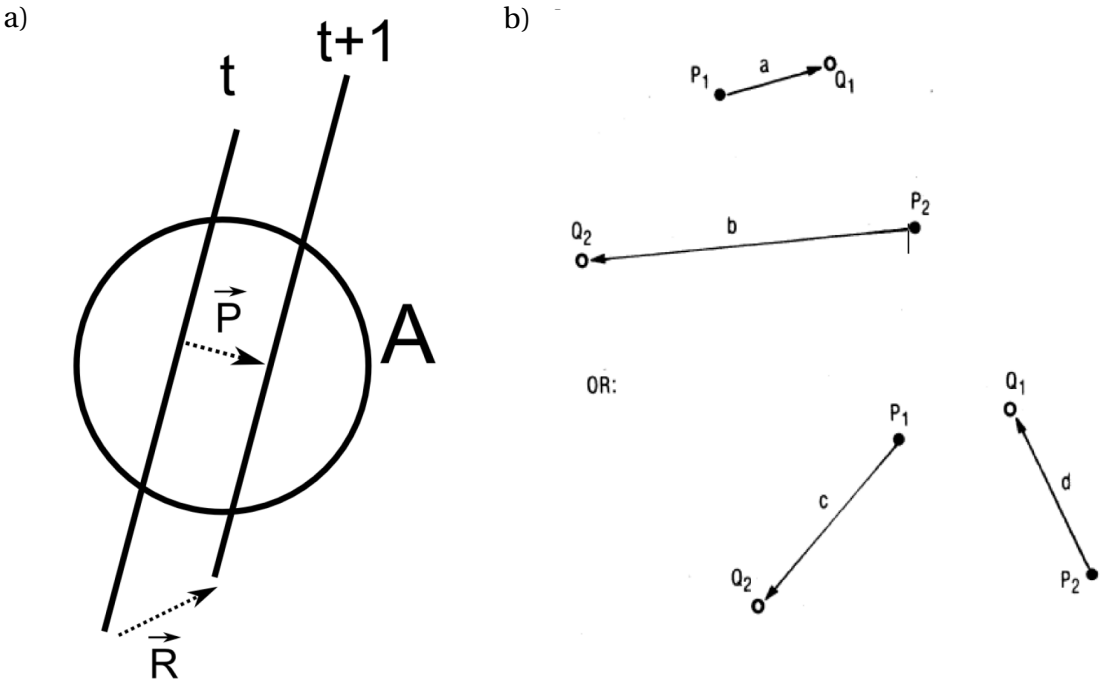

Figure 1.33.: Aperture problem. a) Pin-hole aperture showing 2 features in consecutive time steps, b) The correspondence problem (reprinted from Ullman 1981).

1.5.2. Measuring optic flow

Measuring optic flow sparked several methods for its computing from image sequences. Most of these methods can be classified in two main groups:

- Raw information-based: where the motion measurements are based directly on the local changes in light intensity values of the images
- Feature matching-based: where the images are pre-processed aiming to identify features such as edges, lines or regions and then measure motion by "matching" these features over the images and measuring the changing positions.

The raw information methods are based on the raw information of the pixels. To begin, consider a scalar function that describes the raw light intensity values of an image (I(x, y)), at a coordinate [x,y]. Then, the optic flow can be defined as:

$$\mathcal{O}(x,y) = [\mathbf{e}_x, \mathbf{e}_y]^T \approx [\frac{\Delta x}{\Delta t}, \frac{\Delta y}{\Delta t}]^T \tag{1.16}$$

Eq. (1.16), describes the displacement (x,y) of these intensity values between two subsequent images at a frequency of $1/\Delta t$. By considering that brightness remains nearly constant between two subsequent image frames, ($I(x,y,t) \approx I(x+\Delta x, y+\Delta y, t+\Delta t)$) and a first-order approximation of the Taylor series of the latter, the following continuity equation can be derived:

$$\nabla I\mathcal{O} + \frac{\delta I}{\delta t} = 0 \tag{1.17}$$

where $\nabla = [\frac{\delta}{\delta x}, \frac{\delta}{\delta y}]^T$ is the spatial gradient operator.

Eq. (1.17) has two unknowns ($\mathbf{e}_x, \mathbf{e}_y$) and can not be solved alone, another set of equations are needed. Note, that this is a direct consequence of the aperture problem (Fig. 1.33). There are two approached to find the extra equations, either local or global, commonly known as local and global differential methods.

The most well known and widely accepted local method is the Lucas-Kanade method (Lucas et al. 1981). This method is based on the assumption that the image changes similarly across the whole frame and that the pixel changes are small. In other words, that the velocity vectors $\mathbf{e}_x$ and $\mathbf{e}_y$ are nearly constant within a small area around a pixel. This method divides the images in regions, for which Eq. (1.17) is calculated. Finally, the values of the optic flow field are then computed by performing a least-squares minimization on $\mathcal{O}$. The robustness of this method can be improved by weighting the least-squares solution, giving more importance to the center of the regions. The advantage of local methods is the relatively high robustness under noisy conditions, although if the optic flow is not uniform it can be fairly unreliable.

The most commonly used global method is the Horn-Schunk method (Horn et al. 1981). This method is based on variational calculus, aiming to minimize the function obtained by adding a smoothness constraint. The advantage of global methods is that they allow optic flow to be smoother and denser but with the cost of a higher computational cost.

The feature matching methods to measure the optic flow approximate the vectors by the displacements that best fit some key identifier in the image sequence (Fig. 1.34). These identifiers could be regions, features, patterns, objects, edges surfaces, etc. Usually this is performed by maximizing a similarity measure, like correlation coefficients, such as the Pearson cross-correlation coefficients. To reduce the computational cost, other similarity measures can be utilized, such as the absolute value of differences or the sum of squared differences.

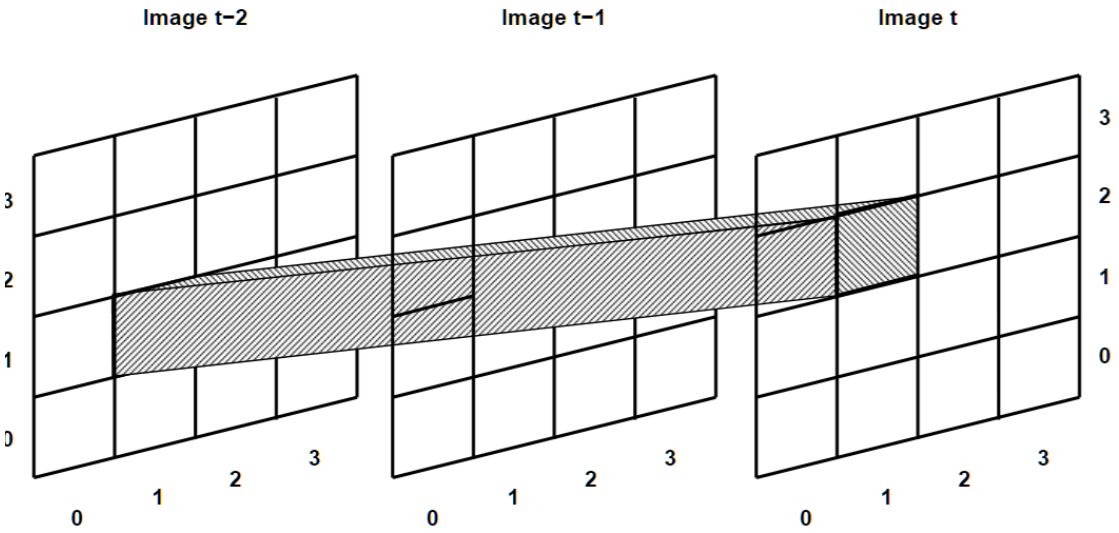

Figure 1.34.: Feature matching between time steps (reprinted from Camus 1995).

Feature matching highly relies on the ability to discern and correlated (match) two different blocks over time. It is highly impacted by false positives, leading to a wrong estimation of the optic flow. This wrong estimation increases when the frame rate is low compared to the feature movement.

To overcome the matching over time issue, Ruffier et al. (2003) has been proposed to have multiple sensors separated in a known angle ($\Delta\Phi$), and evaluate the time between detections of the same feature[8]. Note that the evaluation of time between sensors or, more commonly knows as, "time of travel" (TOT) scheme has been seen directly measured on insects compound eyes (Franceschini et al. 1989). Using TOT measurements, it is possible to directly estimate the time delay Δt between 2 neighboring sensors induced by a visual motion. Then, the calculation of the optic flow becomes $\mathcal{O} = \Delta\Phi/\delta t$.

[8]It can be also defined as a token. In this case, it would be token-matching techniques. For this study, the token, and a feature are the same.

1.6. Vision and Collective motion

Section 1.2 showcased how collective motion is ubiquitously present in nature. Section 1.3 presented how collective motion has been described over the years. Finally, section 1.5 describes the important role of vision and visual information, such as optic flow, on animals. A natural progression is to explain animal collective motion from visual information. To include visual information into collective motion models, the first step was to select the relevant particles using vision (Fig. 1.35). Schilling et al. (2022) present a comparison between the particle selection with 4 different methods, one of them being vision. The overall performance of all methods was found equivalent.

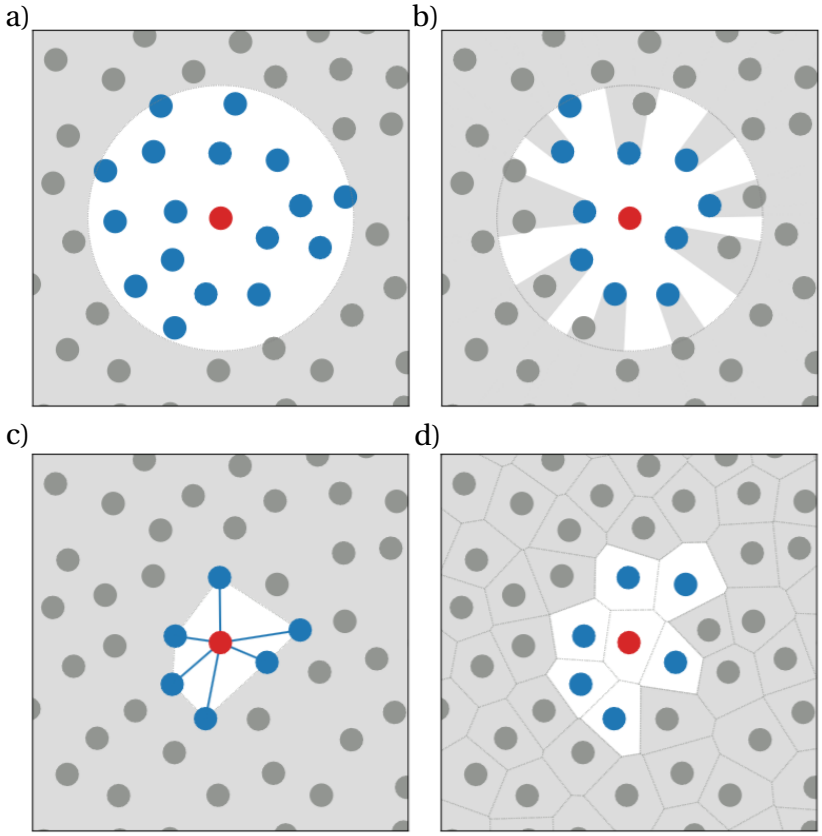

Figure 1.35.: Schematic visualization of different neighbor selection strategies: metric, visual, topological, and voronoi. Perspective of a focal agent within a swarm (central red disk) that selects agents (blue disks) and discard others (gray disks) depending on the following selection criteria: a) all agents within a distance radius, b) all visible agents within a metric radius, assuming agents are equally sized and have an omnidirectional camera at their center, c) only the n closest agents (n = 6 shocased), irrespective of their distance, d) only those agents that belong to a neighboring Voronoi region (reprinted from Schilling et al. 2022).

Soria et al. (2019), presented an extensive analysis of the traditional Boids model using vision, or the blindspot, as limiting factor. The visual blindspot was found to not affect the capabilities of the model to reproduce collective behavior, just shifting where the phases are reproduced. D. J. Pearce et al. (2014) presented a model that included visual information as projections of presence of individuals. The average of the projections on the visual field would generate a heading vector equivalent to an attraction law. The proposed attraction, with the alignment implementation of the Viscek model, was successful on the reproduction of collective motion. Collignon et al. (2016) used the same idea of a projection of the flock to the focal point of each individual to implement a collective motion model based on the retinal probability function of the flock. Using this probability function, a wall description and a von Mises distribution clustered around the individual, they manage to replicate experimental data of a flock of 5 fish.

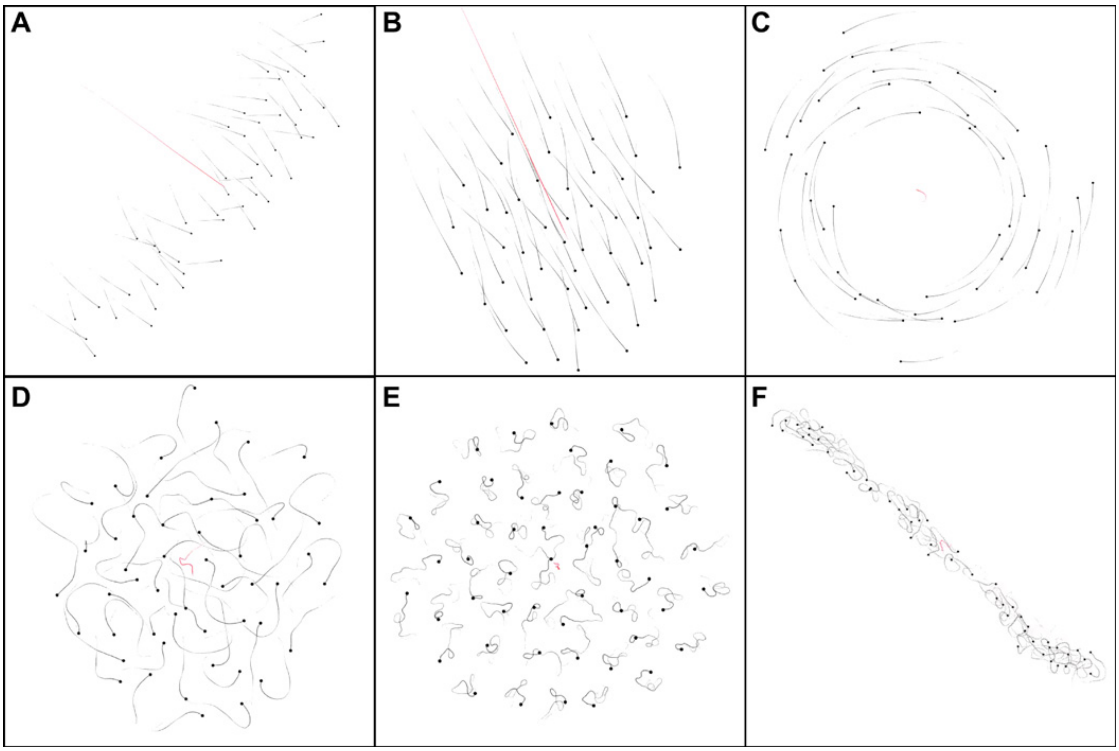

Figure 1.36.: a) Polarized on a line perpendicular to the movement, b) Polarized in a circular shape, c) Rotating. No preferred direction, d) Swarm behavior, e) Crystal-like configuration , f) Tube-like configuration (reprinted from Bastien et al. 2020).

Krongauz et al. (2024), presented a collective motion model for locust. Using the geometric description a locust as a rectangle, they evaluated the possible movement in terms of visual information. Using the movement in visual information terms, a

Viscek model on an area around the individual was able to reproduce schooling on several periodic areas (areas where the boundaries are connected). Bastien et al. (2020) presented a model that used a binary representation of the visual field. The Binary visual field was partially derived in space and time to reproduce various collective motion phases (Fig. 1.36). The derivative controlled the velocity and angle of each of the particles, and the state space of the model was the strength of these derivatives.

The main role vision has played on collective motion, has been as a filter. Most models that include visual information use it as the metric to limit the information available. Another notable use is to complement already established information, such as velocities or headings. Finally, visual information was used to completely redefine collective motion. This work aims to introduce visual information as an independent source of information that aligns with traditional collective motion models.

1.7. Summary

Collective motion is a fascinating natural phenomenon present across all length and time scales. The same emerging patterns are present in a multitude of biological systems regardless of locomotion, size, or sensing capabilities. It is a subject that has sparked years of research, led to incredible advancements, and has furthered our knowledge of nature. And yet, the current state of understanding of collective motion is still based on assumptions and measurements that are not found in nature. In other words, nature has been explained using assumptions that do not hold in nature. **How can collective motion be explained from a biologically plausible point of view?**

Some studies have proposed collective motion models based on the use of visual information, recognizing its key role in nature (Ito et al. 2024; Bastien et al. 2020; Krongauz et al. 2023). However, despite its importance, visual information has primarily been relegated to a secondary source or filter. When collective motion is explained solely from a visual perspective, the resulting interpretations are often so markedly different that they necessitate entirely new frameworks, which cannot easily be correlated with traditional collective motion models. This raises a critical question: **does a visual collective motion model need to disregard traditional perspectives?**

The use of visual information provides tools that are directly correlated with the parameters employed in traditional collective motion models, such as velocities and distances. For instance, optic flow serves as a visual cue that quantifies relative velocities. Moreover, optic flow has been demonstrated to be both sensed and utilized by animals. However, to date, no existing model has fully integrated optic flow to describe collective motion. This omission is primarily attributed to the simplistic assumptions underlying traditional collective motion models. These models predominantly simulate flocks using massless spherical particles governed by the simplest forms of kinematics. In the rare instances where dynamic particle models are included, they are typically constrained to first-order dynamics.

The reductionist approach of self-propelled particle (SPP) models is often employed to accommodate the large numbers required for simulating large flocks in collective

motion. However, these simulations are not practically feasible for robotic implementations due to current technological limitations. Instead, robotic implementations tend to reduce the number of individuals while enhancing the fidelity of interactions between simulation and experimentation. For effective robotic collective motion modeling, it is crucial to balance fidelity with the capability to expand to larger flocks. Achieving this balance depends on a close correspondence between the simulated individuals and their robotic counterparts.

This raises an antagonic set of complementary questions: **How can self-propelled particle simulations be efficiently modified to achieve a better alignment between simulation and robotic experimentation?** or **How can robotic experimentation can be simplified to introduce more individuals and still align with simulation?**

This work aims to answer these questions. It presents a biologically plausible model of collective motion that uses vision and optic flow. Using these optic cues, the visual model follows the traditional formulation of self-propelled particles and is capable of reproducing three phases of collective motion—swarming, schooling, and milling. Moreover, this work describes the implementation of a supplemented visual model in a flock of 10 spherical robots. The implementation and simulation of the visual model reproduce the same phases and show a similar distribution in the state space of the model parameters. The work is organized as follows.

Chapter 2 introduces a model of collective motion constructed from visual information to enforce the traditional rules. This chapter presents the results of this model, a comparison between the visual model and a traditionally formulated model, and finally shows the equivalency between the two. The visual model provides a tool that allows any particle physics or statistical physics-based model to be referred to as a biologically plausible source of information. Furthermore, it shows that it is possible to explain nature on nature's terms within the traditionally constructed physics model.

Chapter 3 presents a robots-in-the-loop real-world implementation of self-propelled particle models using the visual model. This chapter validates a robotic flock of spherical robots against a numerical simulation of non-inertial particles in the frame of collective motion. The chapter introduces a robotic platform that can be used for collective motion research with kinematics similar to those of self-propelled particle models. In other words, it presents a tool that enables translation from a model that assumes massless particles to a commercial robot capable of interacting with the real world and vice versa.

In conclusion, this work describes tools to explain collective motion using biological information without losing the history and lessons of traditional physics. A bridge between the non-inertial world of numerical simulations of self-propelled particles and the world of robotics was built. Finally, an equivalency between spherical robots and collective motion models was established.

2. Reconciling vision and particle physics.

Contents

2.1. Abstract 69
2.2. Introduction 69
2.3. Model 72
2.4. Numerical Simulations 74
2.5. Discussion 76
2.6. Conclusion 77
2.7. Supplementary Information 78
2.7.1. Mathematical development 78
2.7.1.1. Visual integration constant proof 78
2.7.1.2. Alignment on visual terms 79
2.7.1.3. Blind spot 79
2.7.2. Digital implementation 80
2.7.2.1. Visual model 80
2.7.2.2. Algorithm to detect corresponding shades 82
2.7.2.3. Omniscient model 83
2.8. Reconciling vision and particle physics 84

Vision plays a critical role in animal movement, with the eye serving as an organ highly sensitive to both light and motion. Animals rely on their visual system to accomplish a range of tasks, including detecting and using apparent motion to perform complex maneuvers individually. Additionally, animals exhibit a strong tendency to move in groups, a behavior that can be described and replicated based on detailed observations. From these observations, emergent patterns can be discerned within group dynamics. Furthermore, it can be determined which individuals, or groups of individuals, exert the most influence on the group.

However, there remains a gap between our understanding of animal perception and our knowledge of animal motion. The models of animal motion are based on artificial metrics instead of natural cues. To explain this, completely new models linking collective motion and biological perception should be developed. While I believe such an approach is feasible, it may not be essential. In recent work Castro et al. 2024b, a new kind of model was introduced—a model that aims to explain and reproduce collective motion by focusing on animal perception, specifically vision. This new "3A model" can be compared to standard Self-Propelled Particle (SPP) models

and relies solely on visual information to replicate swarming, schooling, and milling behaviors. Notably, a unique bistable state was found and explored in detail.

Modeling collective behaviors from optic flow and retinal cues.

Castro, D., Ruffier, F., & Eloy, C. (2024). "Modeling collective behaviors from optic flow and retinal cues." *Physical Review Research*, 6(2), 023016.

2.1. Abstract

Animal collective behavior is often modeled with self-propelled particles, assuming each individual has “omniscient” knowledge of its neighbors. Yet, neighbors may be hidden from view and we do not know the effect of this information loss. To address this question, we propose a visual model of collective behavior where each particle moves according to bio-plausible visual cues, in particular the optic flow. This visual model successfully reproduces three classical collective behaviors: swarming, schooling, and milling. This model offers a potential solution for controlling artificial swarms visually.

2.2. Introduction

Collective animal behavior is a widespread phenomenon in nature, ranging from the mesmerizing movements of starling murmurations to the coordinated motion of cattle herds Liebchen et al. 2017; Caussin et al. 2014; Barberis et al. 2016; Silverberg et al. 2013. These collective behaviors are commonly modeled with self-propelled particles: individuals with an intrinsic speed orient themselves based on a set of rules, alignment, attraction, and avoidance Vicsek et al. 1995; C. W. Reynolds 1987; Ginelli et al. 2010; Albano 1996. These “3A” rules have been successful in reproducing different collective behaviors, such as swarming (no orientational order), schooling (high orientational order), and milling (the group coordinately swirls in a circular pattern) Katz et al. 2011; Mishra et al. 2010; Solon et al. 2015; Vicsek et al. 2012. Furthermore, numerous studies used these rules to replicate some of these phases on artificial robotic swarms Virágh et al. 2014; Wang et al. 2019; Y. Shen et al. 2022.

In self-propelled particle (SPP) models , the rules of attraction, alignment, and avoidance are typically applied with the simplifying assumption that each individual possess idealized senses, is "omniscient" and gauges perfectly the position, distance, orientation, and velocity of its neighbors Bode et al. 2010; Mateo et al. 2019. However, this assumption may not hold in practice, especially when some neighboring individuals are hidden from view Soria et al. 2019; Davidson et al. 2021; Strandburg-Peshkin et al. 2013; Lemasson et al. 2013; Ballerini et al. 2008b.

The causal link between visual cues and collective behavior has been shown through several vision-based biological models Lemasson et al. 2009; Collignon et al. 2016; Qi et al. 2022; Aoki 1984. As a result, different ways of incorporating vision in SPP models have been suggested. The most widespread approach consists of using vision to filter information Krongauz et al. 2024; Ariel et al. 2014. One of these models suggested that flocks of starlings adjust their density to reach a state of "marginal opacity" D. J. Pearce et al. 2014. However, this density adjustment does not seem to be widespread, as certain animals such as fish can form opaque schools Ballerini et al. 2008a; Ginelli et al. 2010.

In a recent study, Bastien and Romanczuk proposed a model of collective behavior based purely on vision capable of reproducing most of the collective behaviors Bastien et al. 2020. Their model simulates each individual's response to a projection of the visual field, rather than relying on omniscient information. However, the portrayed milling phase is uncoordinated, meaning that the particles turn in both directions in the same swirl. And, while they claim to use the simplest possible equations of movement that satisfy fundamental symmetries, their model involves six parameters that are hard to relate to the classical 3A rules. In addition, four terms of their model involve spatial and temporal derivatives that would use important computing resources. Instead, we propose to use bio-plausible visual cues that can be measured directly by a visual sensor, i.e. the optic flow and the retinal position[1].

Optic flow refers to the apparent angular velocity of objects in the visual field due to the relative motion between the observer and its surroundings. Numerous animal species perceive and use optic flow for a variety of tasks. Insects use it to navigate in crowded settings Portelli et al. 2010, evade ground obstacles, and control their landing Franceschini et al. 2007. Fish use it for navigation Bonnen 2023, and birds use it during takeoff Serres et al. 2019. Optic flow is ubiquitous in nature. It involves specialized neurons, well-identified in invertebrates, which have inspired bio-inspired sensors dedicated to optic flow Ruffier et al. 2005; Expert et al. 2015; Floreano et al. 2013; Schilling et al. 2022. These visual sensors can provide panoramic optic flow sensing Floreano et al. 2013; Expert et al. 2015 and allows for direct measurement and on-board panoramic use of this visual cue Bergantin et al. 2021; De Croon et al. 2022.

In this article, we propose a non-stereoscopic vision-based model of collective behavior, inspired by animal vision. This model is intended to be implementable on autonomous robots equipped with visual sensors, i.e. the robots do not need to communicate with a central command or/and to know their neighbors' relative coordinates, their relative positions between each other or to be georeferenced. Our model aims at bridging the gap between traditional SPP models that rely on omniscient information and biomimetic visual approaches.

[1]The retinal position is the relative location of an object referred to it's position on the visual field. The visual field is centered at the line of sight of the individual.

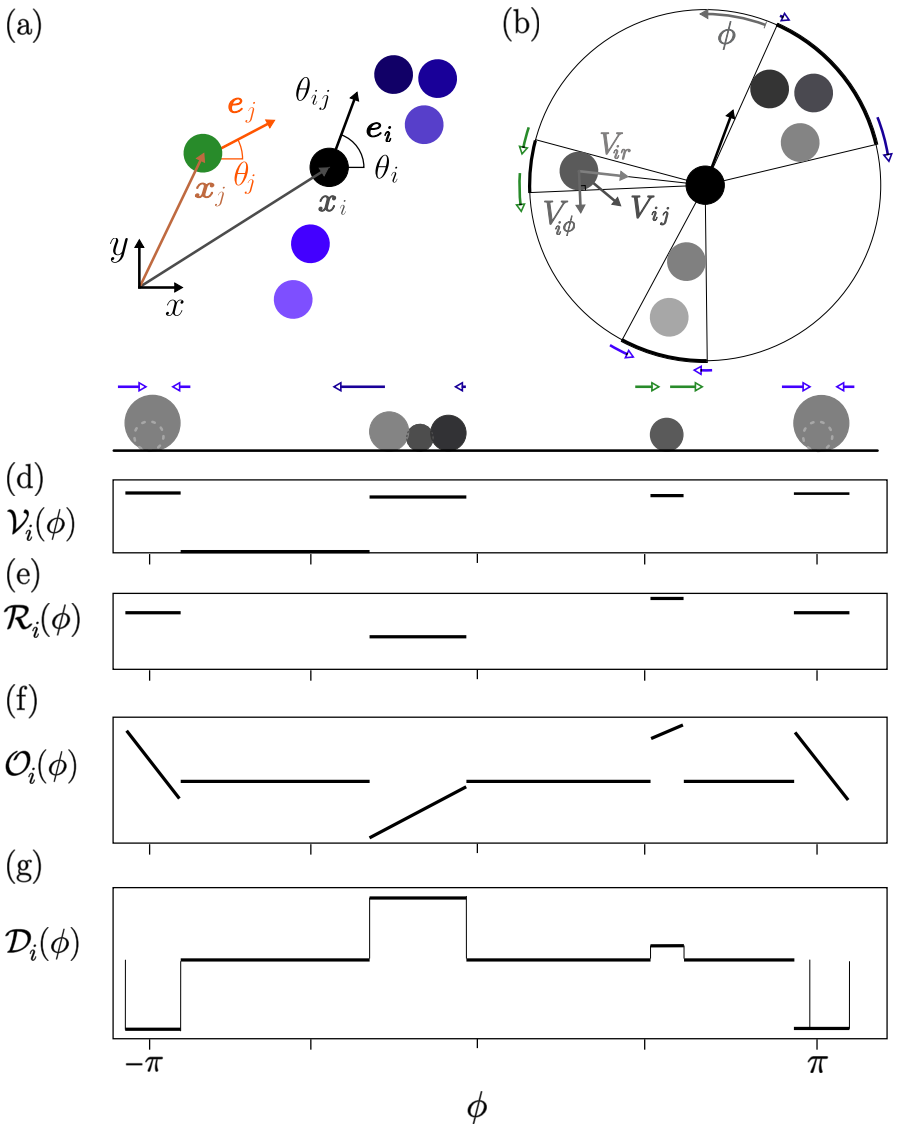

Figure 2.1.: Notations used and principle of the visual cues. (a) The particle i is located in $\boldsymbol{x}_i$ and θ_i is its heading. θ_{ij} is the position of the projection of particle i to the visual field (retinal position) of particle j for the i-th particle. (b) The angle ϕ is the line of sight with respect to the particle's heading. The vector $\boldsymbol{V}_{ij} = U(\boldsymbol{e}_j - \boldsymbol{e}_i)$ is the relative velocity of particle j perceived by i. It can be decomposed into radial and azimuthal components, V_{ir} and $V_{i\phi}$. (c) Point of view of the particle i, the arrows represent the optic flow generated by the relative angular velocity of each particle. (d) The visual field $\mathcal{V}_i(\varphi)$ is a binary representation of (c). It is formed of a set of "shades". (e) The function $\mathcal{R}_i(\phi)$ represents the estimated distance associated to shades, assuming they are associated with a single individual of radius a. (f) Optic flow $\mathcal{O}_i(\phi)$ associated with the apparent angular velocity of the shade edges. (g) Optic flow divergence $\mathcal{D}_i(\phi)$.

2.3. Model

We consider a system of N self-propelled particles in two dimensions. Each particle is a circular object with radius a, moving with a constant speed U. The position of the ith particle is noted by $\boldsymbol{x}_i \in \Re^2$ and its direction $\boldsymbol{e}_i = [\cos\theta_i, \sin\theta_i]$, with θ_i its heading, θ_{ij} would be the subtraction of the heading angle of the particle i and the heading angle of the particle j (Fig. 2.1 a)).

We model the interactions between particles using changes in their angular velocity associated with attraction, alignment, and noise. Specifically, the equations of motion can be written as follows

$$\dot{\boldsymbol{x}}_i = U\boldsymbol{e}_i, \tag{2.1a}$$
$$\dot{\theta}_i = k_\odot \omega_\odot + k_\parallel \omega_\parallel + k_\eta \eta, \tag{2.1b}$$

where dots denote temporal derivatives and $\eta(t)$ is a standard Wiener process representing rotational noise. The functions $\omega_\odot$ and $\omega_\parallel$ are $O(1)$ functions representing attraction and alignment. The parameters $k_\odot$, $k_\parallel$ and k_η control the strength of attraction, alignment, and noise, respectively. For simplicity, we did not include an avoidance rule, as it is not a required rule to reproduce collective behaviors Caussin et al. 2014; Morin et al. 2015; Vicsek et al. 2012.

We begin by introducing an omniscient model that will serve as a reference. This model is inspired by a data-driven fish model Calovi et al. 2014; Filella et al. 2018, with the difference that each particle interacts with all the others. The attraction and alignment terms are given by

$$\omega_\odot^{\text{omni.}} = \left\langle \sum_{j=1}^{N} \|\boldsymbol{x}_j - \boldsymbol{x}_i\| \sin(\theta_{ij}) b_\epsilon(\theta_{ij}) \right\rangle, \tag{2.2a}$$
$$\omega_\parallel^{\text{omni.}} = \left\langle \sum_{j=1}^{N} \frac{(\boldsymbol{e}_i \times \boldsymbol{e}_j)\cdot\hat{\mathbf{z}}}{\|\boldsymbol{x}_j - \boldsymbol{x}_i\|^2} b_\epsilon(\theta_{ij}) \right\rangle, \tag{2.2b}$$

where $b_\epsilon(\phi) = 1 + \epsilon\cos\phi$ models the blind angle. When $\epsilon = 0$, b_ϵ is isotropic; when $\epsilon = 1$, the particle cannot see behind itself (See Section 2.7.1.3). $\hat{\mathbf{z}}$ is the unitary vector that is orthogonal to the $(\hat{\mathbf{x}}, \hat{\mathbf{y}})$-plane of movement of the particles. The brackets denote a normalization defined as

$$\left\langle \sum_j f(j)\sin\theta_j \right\rangle = \sum_j f(j)\sin\theta_j \Big/ \sum_j |f(j)|. \tag{2.3}$$

To model visual perception, we assume that each particle senses a visual field $\mathcal{V}_i(\phi)$, where ϕ represents the angle between the particle's heading and the line of sight (Fig. 2.1 c)). The function output is binary, indicating the presence or absence of a shade in the visual field (Fig. 2.1b)-d)).

Using the information from the visual field $\mathcal{V}_i(\phi)$, we can derive the function $\mathcal{R}_i(\phi) = a/\sin\big(\Delta\phi/2\big)$, where $\Delta\phi$ represents the width angle of a retinal object (Fig. 2.1 e)). With

this definition, when a shade is associated with a single particle, $\mathscr{R}_i(\phi)$ represents its distance from the viewer.

Temporal changes in the features of the visual field can be used to calculate the optic flow $\mathscr{O}_i(\phi)$. A simplified optic flow is estimated by assuming that each shade has a pattern that moves and deforms with it. It results that $\mathscr{O}_i(\phi)$ is a linearly interpolated function of the angular velocity between two features of the shade, its rising and falling edges [Fig. 2.1(f)]. This method to compute the optic flow computation corresponds to a cross-correlation of visual field features Ullman 1981, which are known to occur in animal eyes Franceschini et al. 1989; Eichner et al. 2011. Similarly, we can compute the optic flow divergence $\mathscr{D}_i(\phi)$ by derivating the optic flow $\mathscr{O}_i(\phi)$ [Fig. 2.1(f)]. Due to the piece-wise linear nature of $\mathscr{O}_i(\phi)$, $\mathscr{D}_i(\phi)$ is a piece-wise constant function.

The optic functions $\mathscr{V}$, $\mathscr{R}$, $\mathscr{O}$, and $\mathscr{D}$ are inspired by animal vision. These functions can easily be computed by a man-made vision system. We will now use these functions to the attraction and alignment terms of a vision model.

In this visual model, the attraction and alignment terms are given by

$$\omega_{\odot}^{\text{visu.}} = \left\langle \int_{-\pi}^{\pi} \mathscr{R}_i^2(\phi)\, b_\epsilon(\phi) \sin\phi \, d\phi \right\rangle, \tag{2.4a}$$

$$\omega_{\parallel}^{\text{visu.}} = \left\langle \int_{-\pi}^{\pi} \frac{(\boldsymbol{e}_i \times \boldsymbol{V}_{ij}) \cdot \hat{\mathbf{z}}}{U \mathscr{R}_i(\phi)} b_\epsilon(\phi)\, d\phi. \right\rangle, \tag{2.4b}$$

where the brackets denote the normalization given by Eq. (2.3) with the sum replaced by the integral. The attraction and alignment terms of the visual model are constructed to be similar to those of the omniscient model given in Eqs. (2.2a)-(2.2b). The difference in the exponent of $\mathscr{R}$ comes from the additional $\Delta\phi \sim 1/\mathscr{R}$ arising from the integration.

In the alignment term, $\omega_{\parallel}^{\text{visu.}}$, the cross-product $\boldsymbol{e}_i \times \boldsymbol{e}_j$ is evaluated from the visual information. This is done by using the optic flow $\mathscr{O}_i(\phi)$ and its divergence $\mathscr{D}_i(\phi)$, which are related to the velocity of particle j with respect to particle i (Fig. 2.1 b)). Specifically, the radial component is given by $V_{ir} = -\mathscr{R}_i(\phi)\, \mathscr{D}_i(\phi)$, and the azimuthal component by $V_{i\phi} = \mathscr{R}_i(\phi)\, \mathscr{O}_i(\phi)$. We can use these components to calculate the vector $\boldsymbol{V}_{ij}$ in polar coordinates (r,ϕ) as $\boldsymbol{V}_{ij} = (-\mathscr{D}_i, \mathscr{O}_i)\mathscr{R}_i/U$, while $\boldsymbol{e}_i$ can be expressed as $(\cos\phi, -\sin\phi)$. It results that

$$\frac{\boldsymbol{e}_i \times \boldsymbol{V}_{ij}}{U\mathscr{R}_i(\phi)} \cdot \hat{\mathbf{z}} = \frac{-\mathscr{D}_i(\phi)\sin\phi + \mathscr{O}_i(\phi)\cos\phi}{U}. \tag{2.5}$$

The alignment term is thus the sum of two terms: one proportional to the derotated optic flow $\mathscr{O}$ and sensitive to the azimuthal velocity of neighbors, and one proportional to the optic flow divergence $\mathscr{D}$ and sensitive to the radial velocity. When computing $\boldsymbol{e}_i \times \boldsymbol{V}_{ij}$ with Eq. (2.5), we remove the particle rotation with angular velocity $\dot{\theta}_i$ from the optic flow $\mathscr{O}_i(\phi)$.

The equations of motion presented in Eq. (2.1), along with the attraction and alignment terms derived in Eqs. (2.4a) and (2.4b), provide a model of collective behavior based on realistic visual cues. To make the problem dimensionless, we chose $a = 1$

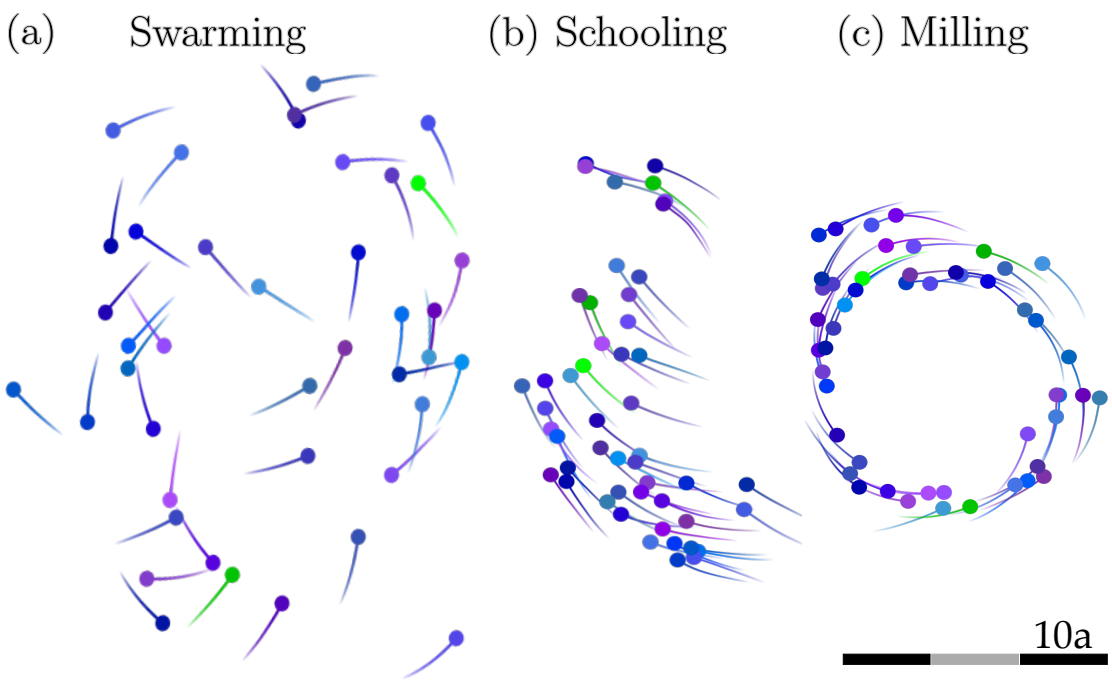

Figure 2.2.: Illustration of the phases observed for $N = 50$ individuals. The scale bar measures $30a$ in total ($10a$ for each section). The parameters $\epsilon = 1$ and $k_\eta = 0.01$ are fixed. We observe three phases when varying the other two parameters: (a) swarming ($k_\odot = 0.1$, $k_\parallel = 0$); (b) schooling ($k_\odot = 0.06$, $k_\parallel = 0.2$); and (c) milling ($k_\odot = 0.1$, $k_\parallel = 0.04$).

. . .

and $U = 1$. With this approach, four dimensionless parameters remain, the strengths of noise attraction and alignment, k_η, $k_\odot$, $k_\parallel$, and the blind angle parameter ϵ.

2.4. Numerical Simulations

To explore the effect of these parameters on collective behaviors, we performed numerical simulations with $N = 50$ particles. Initially, the particles are randomly placed in a square of side aN with random headings (See Sec. 2.7.2). The dynamical system described by Eqs. (2.1) and (2.1) is solved numerically using a discrete implementation of Eqs. (2.4a) and (2.4a) (See 2.7.2). We examined the effect of the time step δt by conducting simulations with $\delta t = 0.001$, 0.01, and 0.1 (See Sec A.2.1). However, no significant differences were observed on the collective behavior, and a time step of $\delta t = 0.1$ was selected for the remaining simulations to ensure computational efficiency.

We first set $k_\eta = 0.01$ and $\epsilon = 1$ (maximum blind angle) and explore the effects of the two remaining parameters $k_\odot$ and $k_\parallel$ in the visual model. Our simulations show three distinct dynamical phases (Fig. 2.2 for the visual model, Fig. A.7 for the omniscient model and Sec. A.3.1 snapshots for the evolution of the stable phases seen

on Videos S7-S9 and S10-S12). If the alignment is zero, a disordered swarming phase is observed, where individuals form a group without a preferred direction (Fig. 2.2 a)). When the alignment strength increases, particles begin to align in the same direction, resulting in the schooling phase (Fig. 2.2 b)). If the ratio between the alignment and the attraction strengths is around 0.4, the group exhibits a milling phase (Fig. 2.2 c)), creating a vortex. These three phases (swarming, schooling, and milling) have regularly been observed in (omniscient) self-propelled-particle models Calovi et al. 2014; Filella et al. 2018; Charlesworth et al. 2019; Lu et al. 2022; Shang et al. 2014

To quantitatively distinguish between the different dynamical phases, we introduce three global order metrics: polarization P, milling M, and opacity O Calovi et al. 2014; Filella et al. 2018. These metrics are defined as follows,

$$P = \|\overline{\boldsymbol{e}_i}\|, \tag{2.6a}$$

$$M = \|\overline{\boldsymbol{y}_i \times \boldsymbol{e}_i}\|, \tag{2.6b}$$

$$O = \frac{1}{2\pi}\overline{\int_{-\pi}^{\pi} \mathcal{V}_i(\phi)\, d\phi}, \tag{2.6c}$$

where the overbar represents an average over all individuals and the unit vector $\boldsymbol{y}_i = (\boldsymbol{x}_i - \overline{\boldsymbol{x}_i})/\|\boldsymbol{x}_i - \overline{\boldsymbol{x}_i}\|$ points towards particle i from the center of mass. All three metrics range in the interval $[0,1]$. The polarization P measures the alignment: $P = 0$ corresponds to particles pointing in all directions, $P = 1$ corresponds to a perfectly aligned school. The milling M represents the normalized angular momentum: straight-line formation gives $M = 0$ and perfect milling gives $M = 1$. The opacity O measures the "occupancy" of the visual fields: $O = 0$ when there is no object in the visual field, and $O = 1$ when the entire visual field is obscured.

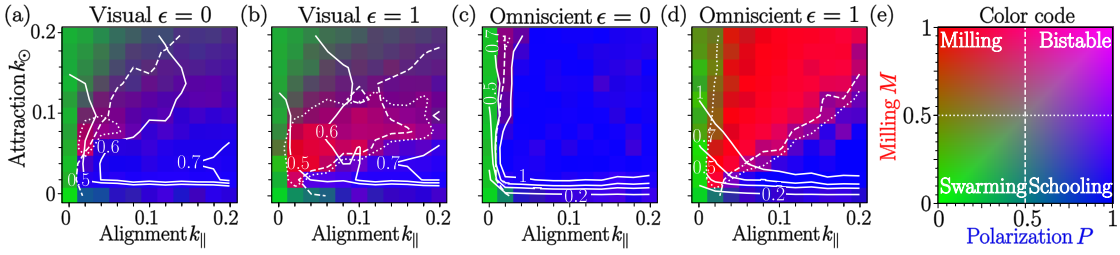

Figure 2.3.: Phase diagrams for $N = 50$ individuals and noise strength $k_\eta = 0.01$. These phase diagrams compare the visual model (a), (b) and the omniscient model (c), (d) for $\epsilon = 0$ (a), (c) and $\epsilon = 1$ (b), (d). The colors represent different values of P and M as shown in (e). The contours show the values of opacity O (white solid), the line $M = 0.5$ (white dotted), and $P = 0.5$ (white dashed).

We now compare the visual and omniscient models by setting the value of the noise to $k_\eta = 0.01$ and exploring the parameter space $(k_\odot, k_\parallel) \in [0, 0.2] \times [0, 0.2]$ for two values of the blind angle parameter ($\epsilon = 0$ or 1). For each parameter set, we ran 10 simulations over long durations ($\Delta t = 5000$). The mean values of P, M, and O were determined by

averaging over the last 1000 time units to ensure that the transient has no influence. The outcomes of these simulations are synthesized in phase diagrams (Fig. 2.3, and expanded on each metric independently on Sec. A.1.1).

2.5. Discussion

Somewhat arbitrarily, we chose to identify the collective phases from the values of the polarization P and the milling M parameters: schooling when $P > 0.5$ and $M < 0.5$; milling when $P < 0.5$ and $M > 0.5$; swarming when $P < 0.5$ and $M < 0.5$; and bistable when $P > 0.5$ and $M > 0.5$ (we will come back to this particular phase below).

Our vision model qualitatively reproduces the phases observed in the omniscient model [Fig. 2.3]. Specifically, we observe the three phases in the vision model: schooling when $k_\odot \lesssim 0.5 k_\parallel$; swarming when $k_\parallel \lesssim 0.3 k_\odot$; and milling or bistability otherwise. When the blind angle parameter increases, it tends to stabilize the milling phase both in the visual model and the omniscient model, as observed in the literature Calovi et al. 2014.

The difference between the two models increases at high values of the alignment and attraction strengths when the opacity O is maximum. In the visual model, the opacity does not exceed 0.7, regardless of the values of ϵ and k_η. In the omniscient model, however, the opacity exceeds 0.8 and even reaches one (Fig. 2.3 c) and d)). This is because the radius a does not play any role in the omniscient model and the characteristic length is set by k_η^2. In both models, as expected, larger noise strength causes a decrease in average opacity (Figs. **??** and A.8).

Now, let us examine the bistable phase. Figure 2.4 shows this phase in the visual model for $\epsilon = 1$, $k_\odot = 0.1, k_\parallel = 0.2$, and $k_\eta = 0.01$. After a transient, the group forms a milling phase until $t \approx 6000$, but eventually it transitions to a schooling phase for $6000 \lesssim t \lesssim 7500$ before returning to a milling phase again, and so on. These transitions show that the system exhibits a noise-induced intermittency between two stable states: milling and schooling. The schooling phase far from the transition resembles a front of parallel individuals (Fig. 2.4 f)). Just before the transition to milling, some individuals move ahead, (Fig. 2.4 c)). Reciprocally, the milling phase just before the transition to schooling opens up, generating a C shape [(Fig. 2.4(e))]. These transitions are fairly stereotyped as they tend to follow the same path in the (M, P) plane (Fig. 2.4 b)). The existence of these intermittent transitions mediated by noise is likely due to a second-order transition between the milling and schooling phases as seen on the metric distribution along the boundaries of both phases (See Sec. A.1.1). This multi-stability has been present on similar 3A models Calovi et al. 2014.

We changed the group size from $N = 5$ to 300 (Figs. A.2 and A.9). Although the group size does not seem to impact qualitatively the phase diagram, small groups tend to favor schooling, whereas large groups tend to favor swarming.

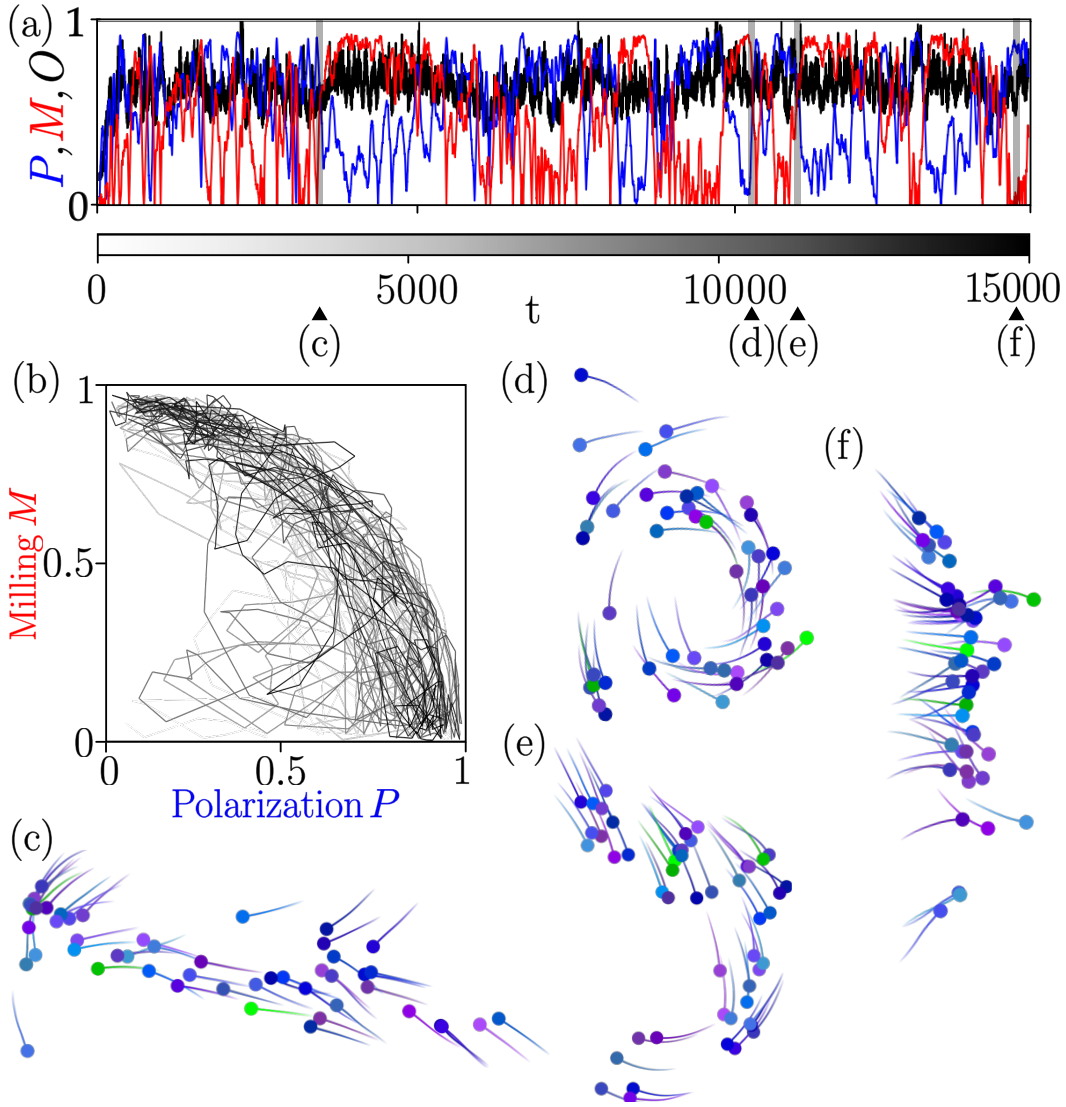

Figure 2.4.: Example of the bistability observed in the visual model (parameter values: $\epsilon = 1$, $k_\odot = 0.1, k_\parallel = 0.2$, and $k_\eta = 0.01$). (a) Time series of the three metrics P, M, and O. (b) Trajectories of the collective behavior in the (M, P)-plane. Illustration of the phases: (c) schooling-milling transition; (d) milling; (e) milling-schooling transition; and (f) schooling.

2.6. Conclusion

In conclusion, we proposed a model based on biologically plausible visual cues. This model successfully reproduces the three classical phases of animal collective behavior: swarming, schooling, and milling. These findings show that visual cues provide enough information to enable collective behavior.

Furthermore, our findings imply potential practical uses for synchronizing groups of

artificial drones, which may be governed by analogous visual stimuli. In future studies, we aim to investigate these opportunities more thoroughly and enhance our model for a more accurate depiction of animal collective behavior in a three-dimensional space.

2.7. Supplementary Information

2.7.1. Mathematical development

2.7.1.1. Visual integration constant proof

The inclusion of an integral over the whole visual field, was stated that introduces a distance inversely proportional term. Here we present the integration of the base terms to proof this statement. We start by the integration of the base attraction found on partially omniscient models:

$$\omega_{\odot} = \int_{-\pi}^{\pi} \mathscr{R}(\phi)\, b_{\epsilon}(\phi) \sin\phi \, d\phi, \tag{2.7}$$

where $b_{\epsilon}(\phi) = 1 + \epsilon\cos\phi$. Due to the piece-wise constant form of $\mathscr{V}(\phi)$, we can rewrite Eq. 2.7 to reflect only the sum of the shades. And each shade k (with $k \leq N-1$) will be described by a rising edge $\varphi_{\Uparrow k}$ and a falling edge $\varphi_{\Downarrow k}$. From these quantities, we compute the edge mid-point $\varphi_k = (\varphi_{\Uparrow k} + \varphi_{\Downarrow k})/2$, and the edge half-width $\Delta\varphi_k = (\varphi_{\Downarrow k} - \varphi_{\Uparrow k})/2$. In this representation $\mathscr{R}(\phi)$ becomes only dependent on $\Delta\varphi_k$, allowing us to remove it from the integral, resulting in:

$$\omega_{\odot} = \sum_k \mathscr{R}(\Delta\varphi_k) \int_{\varphi_{\Uparrow k}}^{\varphi_{\Downarrow k}} b_{\epsilon}(\phi) \sin\phi \, d\phi. \tag{2.8}$$

$$\omega_{\odot} = \sum_k 2\mathscr{R}(\Delta\varphi_k) \sin\big(\Delta\varphi_k\big) \sin\big(\varphi_k\big) \big(1 + \epsilon\cos\varphi_k \cos\Delta\varphi_k\big) \tag{2.9}$$

And here $\mathscr{R}(\Delta\varphi_k) = \frac{a}{\sin\Delta\varphi_k}$, Resulting in:

$$\omega_{\odot} = \sum_k 2\frac{a\sin\big(\Delta\varphi_k\big)}{\sin\big(\Delta\varphi_k\big)} \sin\big(\varphi_k\big) \big(1 + \epsilon\cos\varphi_k \cos\Delta\varphi_k\big) \tag{2.10}$$

$$\omega_{\odot} = \sum_k 2a\sin\big(\varphi_k\big) \big(1 + \epsilon\cos\varphi_k \cos\Delta\varphi_k\big) \tag{2.11}$$

Which loses the distance relationship present on the omniscient model:

$$\omega_{\odot} = \sum_{j\neq i}^{N} ||\boldsymbol{x}_j - \boldsymbol{x}_i|| \sin(\theta_{ij})\, b_{\epsilon}(\theta_{ij}) \tag{2.12}$$

Resulting in Eq. (2.2) to maintain the parallel between both models.

2.7.1.2. Alignment on visual terms

The velocity unit vectors of the j particle perceived by the i-th particle can be expressed on polar coordinates as:

$$\begin{aligned} \boldsymbol{e}_i^p &= [\cos\phi, -\sin\phi] \\ \boldsymbol{e}_j^p &= [-\mathscr{D}(\phi), \mathscr{O}(\phi)]\mathscr{R}(\phi) \end{aligned} \tag{2.13}$$

Note that this notation already contemplates that is self refered information based on the current heading of the i-th particle. Then $\boldsymbol{e}_i \times \boldsymbol{e}_j$ would be:

$$\boldsymbol{e}_i \times \boldsymbol{e}_j = \|\boldsymbol{e}_i\|\|\boldsymbol{e}_j\| \sin\phi \tag{2.14}$$

Here $\boldsymbol{e}_i$ is a unit vector, and $\|\boldsymbol{e}_j\|$

$$\|\boldsymbol{e}_j\| = \sqrt{\mathscr{O}^2(\phi_j) + \mathscr{D}(\phi_j)}\mathscr{R}^2(\Delta\phi_j) \tag{2.15}$$

Then the integral becomes:

$$\omega_\parallel = \int_{-\pi}^{\pi} \boldsymbol{e}_i \times \boldsymbol{e}_j b_\epsilon(\phi)\, d\phi \tag{2.16}$$

$$\omega_\parallel = \sum_k \sqrt{\mathscr{O}^2(\varphi_j) + \mathscr{D}^2(\varphi_j)}\mathscr{R}(\Delta\varphi_j) \int_{\varphi_{\Uparrow k}}^{\varphi_{\Downarrow k}} b_\epsilon(\phi)\sin\phi \tag{2.17}$$

As the optic flow, divergence of the optic flow and the self refered distance are functions that do not depend on ϕ these can be left outside of the integral and result on a similar equation to Eq. 2.8. But instead of having a term depending on distance, we would have a term depending on velocity if it weren't for the $\mathscr{R}(\varphi_j)$ term. That is why to keep the same idea of attraction we include the division of the cross product by $\mathscr{R}$. Note that after integrating Eq.(B.6), we would add an extra $\mathscr{R}(\varphi_j)$ on the denominator as proven on the previous section. Resulting in

$$\begin{aligned} \omega_\parallel = \textstyle\sum_k \sin(\varphi_k) \frac{\sqrt{\mathscr{O}^2(\varphi_k)+\mathscr{D}^2(\varphi_k)}}{\mathscr{R}(\varphi_k)} \\ \left(1 + \epsilon\cos\varphi_k \cos\Delta\varphi_k\right) \end{aligned} \tag{2.18}$$

2.7.1.3. Blind spot

Fig. 2.5 illustrates the function $b_\epsilon(\varphi)$ used to model a blind spot in the visual field. When ϵ is equal to 0 there is no effect and with 1 the effect is maximum.

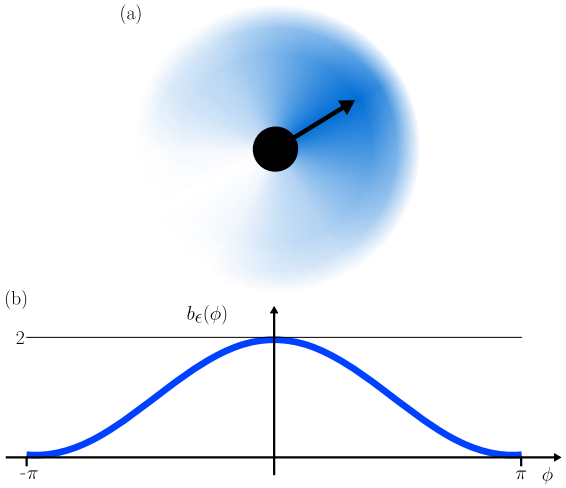

Figure 2.5.: Effect of a blind spot. (a) 2D view: When $\epsilon = 1$, each particle cannot see at its rear (transparent), but the view is normal in front (blue). (b) Function $b_\epsilon(\phi) = 1 + \epsilon \cos\phi$ when $\epsilon = 1$

2.7.2. Digital implementation

2.7.2.1. Visual model

On the main text, we described the continuous form of the equations for attraction Eq. (2.4a) and for alignment Eq. (2.4b). On the supplemental material we expanded the hypotheses we had and developed the continuous form to include the piece-wise linear and piece-wise constant nature of the visual representation (Eqs. (2.11) and (2.18)). Here we go in more depth on how the visual components are derived and calculated, as well as the implemented equations for the numerical simulation.

Consider the point of view of the i-th particle in 2 consecutive time steps for the j-th particle:

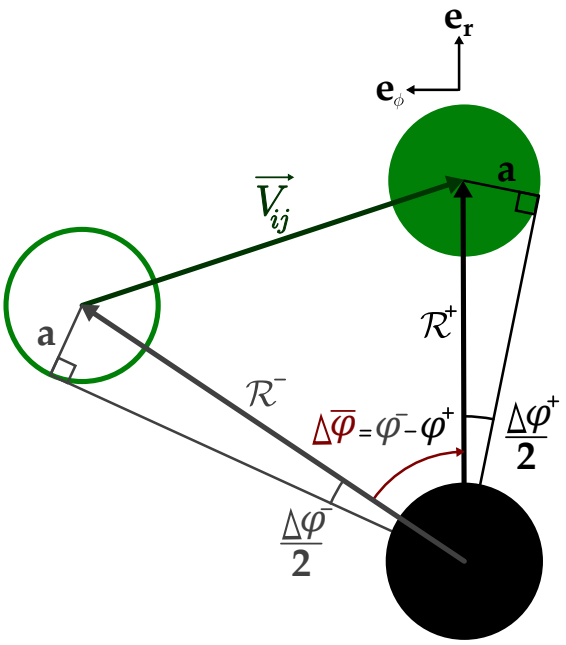

Figure 2.6.: Geometrical equivalent of the optic flow calculation. Point of view of the i-th particle (black) of the j-th particle in $t-1$ (contour) and t (solid color)

Each particle has a radius a, the over index represent + time t and − time $t - dt$. $\boldsymbol{e}_r$ and $\boldsymbol{e}_\phi$ are the unit vectors of the polar coordinates refered to θ_i, where the former is radial and the later azimuthal. $\overrightarrow{V}_{ij}$ is the relative velocity vector perceived by the particle i of the net velocity of the particle j. Then from the geometry of the problem we have

$$\mathscr{R}^{\pm} = \frac{a}{\sin\Delta\varphi^{\pm}} \tag{2.19}$$

Then the $\mathscr{R}$ vectors can be described as follows:

$$\overrightarrow{\mathscr{R}^+} = \mathscr{R}^+ \boldsymbol{e}_r \tag{2.20}$$

$$\overrightarrow{\mathscr{R}^-} = \mathscr{R}^- [\cos\Delta\bar{\varphi}\boldsymbol{e}_r - \sin\Delta\bar{\varphi}\boldsymbol{e}_\phi] \tag{2.21}$$

Consequently the distance associated with the movement of $\overrightarrow{V}_{ij}$ is:

$$\overrightarrow{V}_{ij}\delta t = (\mathscr{R}^+ - \mathscr{R}^- \cos\Delta\bar{\varphi})\boldsymbol{e}_r + \mathscr{R}^- \sin\Delta\bar{\varphi}\boldsymbol{e}_\phi \tag{2.22}$$

$$\overrightarrow{V}_{ij} = \frac{[\mathscr{R}^+ - \mathscr{R}^- \cos\Delta\bar{\varphi}, \mathscr{R}^- \sin\Delta\bar{\varphi}]}{\delta t} \tag{2.23}$$

From Eq. (B.12) we can do the depomposition in radial and azimutal components to calculate $\mathscr{O}$ and $\mathscr{D}$

$$\mathscr{O} = \frac{1}{\delta t} - \frac{\mathscr{R}^- \cos\Delta\bar{\varphi}}{\mathscr{R}^+ \delta t} \tag{2.24}$$

$$\mathscr{D} = \frac{\mathscr{R}^{-} \sin \Delta\bar{\varphi}}{\mathscr{R}^{+} \delta t} \tag{2.25}$$

For the implementation of the visual model alignment, we use Eq. (B.12) to have the cross product of the relative vectors over $\mathscr{R}^2(\Delta\varphi_k)$

$$\omega_{\parallel} = \left\langle \sum_k \sin^2(\Delta\varphi_k) \frac{\vec{V}_{ij} \times \boldsymbol{e}_i^p}{a^2} \left(1 + \epsilon \cos\varphi_k\right) \right\rangle \tag{2.26}$$

As for attraction, we start at Eq. (2.11) and add the Missing $\mathscr{R}$ to have it equivalent to Eq. (2.18).

$$\omega_{\odot} = \left\langle \sum_k 2a^2 \frac{\sin(\varphi_k)}{\sin(\Delta\varphi_k)} \left(1 + \epsilon \cos\varphi_k \cos\Delta\varphi_k\right) \right\rangle \tag{2.27}$$

2.7.2.2. Algorithm to detect corresponding shades

To use Eqs. (2.24) and (2.25), we need to associate shades at two different time steps. But between two time steps, shades can split or join. To detect these events, we employ a minimum distance error to correlate two shades in two consecutive time steps.

In Fig. 2.7, we illustrate the algorithm used, where yellow boxes are associated to corresponding shades. To correspond to each other, shades should be associated to radial and azimuthal velocities bounded by U.

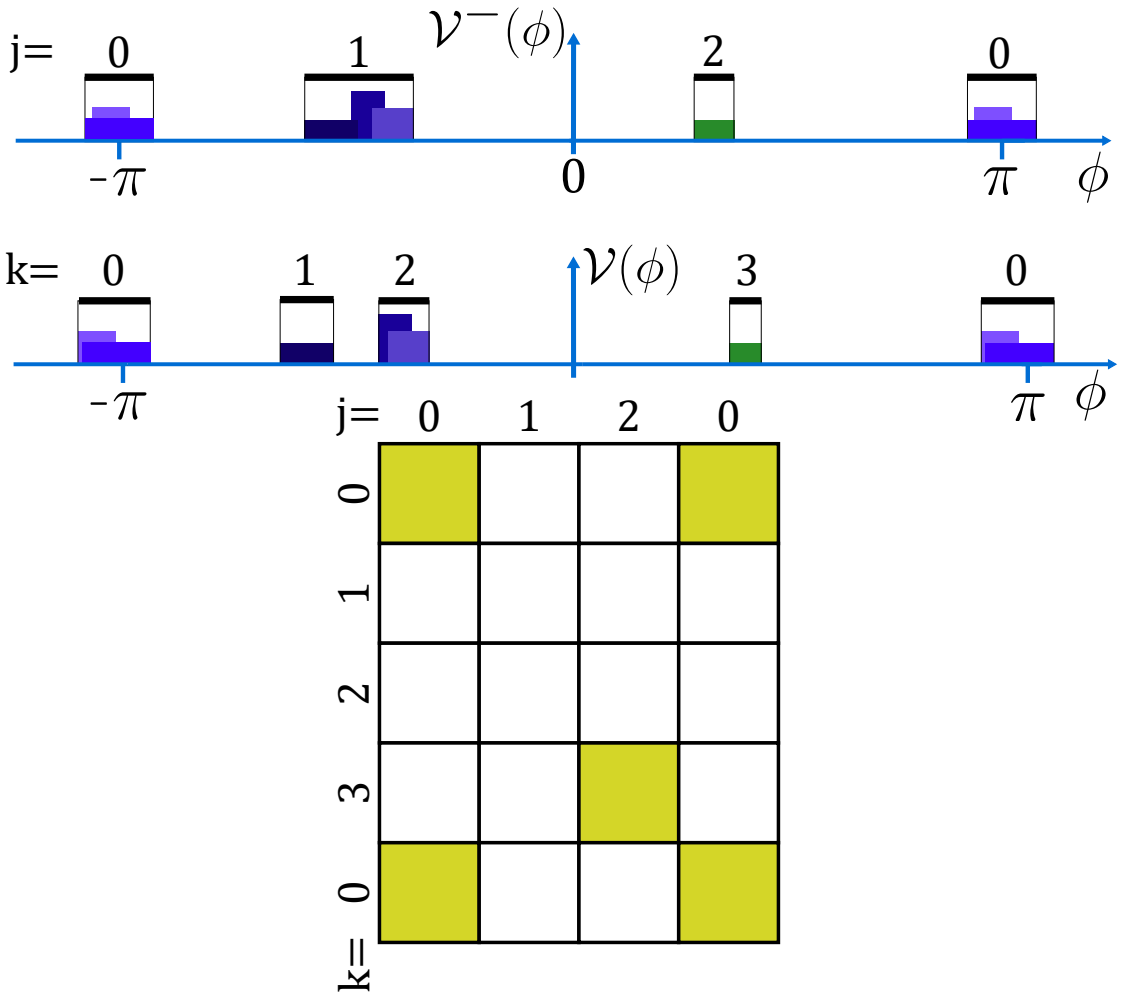

Figure 2.7.: Example of two consecutive visual fields. The matrix at the bottom illustrate the algorithm we use to find corresponding shades between two time steps.

This algorithm is also illustrated on the following pseudo-code:

The model is implemented on python 3.10 (it will be available for download on Castro et al. 2024c. This includes the implementation of the algorithm described by Fig. 2.7, all model equations, data management and figure creation. There is a markdown file that only include the model implementation without data logging.

2.7.2.3. Omniscient model

As previously stated, the model that we propose has an omniscient implementation. Both attraction and alignment started from omniscient parallels to the proposed vision on Eq. (2.4a) and Eq.!(2.4b). Alignment shown on Eq. (2.18) can be converted to omniscient form by taking the parenthesis as b_e and $\sqrt{\mathscr{O}^2(\varphi_k)+\mathscr{D}^2(\varphi_k)}\sin\varphi_k$ as the cross product over the distance function:

Algorithm 1 Corresponding shades between time steps

$\overline{\mathcal{V}_{t-1}} \leftarrow [(\varphi_{\Uparrow l} + \varphi_{\Downarrow l})/2]$ at t-1 $\forall\varphi$ in $\phi_{i,t-1}$
$\Delta\mathcal{V}_{t-1} \leftarrow [\varphi_{\Downarrow l} - \varphi_{\Uparrow l}]$ at t-1 $\forall\varphi$ in $\phi_{i,t-1}$
$\overline{\mathcal{V}_t} \leftarrow [(\varphi_{\Uparrow k} + \varphi_{\Downarrow k})/2]$ at t $\forall\varphi$ in $\phi_{i,t}$
$\Delta\mathcal{V}_t \leftarrow [\varphi_{\Downarrow k} - \varphi_{\Uparrow k}]$ at t $\forall\varphi$ in $\phi_{i,t}$
$\omega \leftarrow \omega_{k,t-1}\delta t$ ▷ Previous known rotation
Require: $l\&k \geq 0$
$VisualDist \leftarrow \|\overline{\mathcal{V}_{t-1}} - \overline{\mathcal{V}_t}\|$ ▷ A k x l matrix
$RefDist \leftarrow min(\sin\Delta\mathcal{V}_t, \sin\Delta\mathcal{V}_{t-1})2U\delta t/a + \omega$
$VisualArea \leftarrow \|\frac{1}{\Delta\mathcal{V}_t} - \frac{1}{\sin\Delta\mathcal{V}_{t-1}}\|$ ▷ A k x l matrix
$RefArea \leftarrow 2U\delta t/a$
for all values in $VisualArea, VisualDist, RefDist$ & $RefArea$ **do**
  if $ValueVisualArea < RefDist$ & $ValueVisualArea < ValueVisualArea$ **then**
    $Pair_t \leftarrow [\varphi_{\Uparrow k}, \varphi_{\Downarrow k}]$
    $Pair_{t-1} \leftarrow [\varphi_{\Uparrow l}, \varphi_{\Downarrow l}]$
  end if
end for

$$\omega_{\parallel} = \left\langle \sum_{j\neq i}^{N} \frac{\boldsymbol{e}_i \times \boldsymbol{e}_j}{||\boldsymbol{x}_j - \boldsymbol{x}_i||^2} b_\epsilon(\theta_{ij}) \right\rangle \tag{2.28}$$

As $\boldsymbol{e}_i$ and $\boldsymbol{e}_j$ are unit vectors then, $\|\boldsymbol{e}_i \times \boldsymbol{e}_j\| = \sin\theta_{ij}$. By consequence, the normalization seen on Eq. (2.3) would work exactly the same.

For attraction the equation would be:

$$\omega_{\odot} = \left\langle \sum_{j\neq i}^{N} ||\boldsymbol{x}_j - \boldsymbol{x}_i|| \sin(\theta_{ij}) b_\epsilon(\theta_{ij}) \right\rangle \tag{2.29}$$

2.8. Reconciling vision and particle physics

The proposed model aimed to relate visual relationships to concepts of traditional particle physics, mainly distance and velocity. The relationships may seem arbitrarily selected, but were the result of an extensive search for the best visual cues that correlate most closely to traditional concepts. In the case of alignment, there is a direct visual cue that relates velocity and vision: optic flow. Optic flow is a relative velocity, but the alignment behavior is always described as a local behavior. Furthermore, in a completely aligned flock, as stated by C. W. Reynolds (1987), the velocity differences and, by extension, the optic flow would be 0. Attraction was an entirely different story. Fig. 2.8 shows the exact same parameters of low alignment and high attraction; it even has similar global metrics, the same scale, and the same evolution. But the particle interaction is fundamentally different.

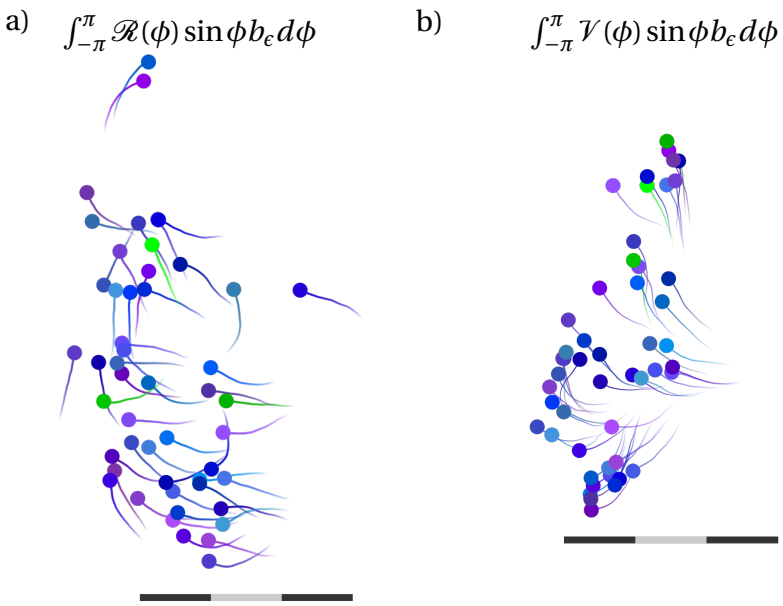

Figure 2.8.: Visual cue for attraction behavior
Phase reproduction for the same parameters but different visual cue to enforce the attraction rule. a) Visual distance, b) Visual size.

Fig. 2.8 a) presents visual attraction based on optic distance (for j particle $a/\sin\Delta\Phi_j$). In the stable state, the particles are uniformly distributed in the area around 15 times the radius of the particles (a), and the flock does not present sub-group attraction or group pairing. It is worth mentioning that during the evolution of the simulations, some rogue particles tend to go 10 to 15a beyond the flock but all return at some point.

Fig. 2.8 b) presents visual attraction based on optic size (for j particle $\sin\Delta\Phi_j$). In this case, the stable state of the flock maintains a slightly higher density (50 particles in a circular area of an approximate radius of 12a), but there is some grouping of 5 to 10 particles. Upon closer examination of these groupings, there is significantly more particle overlap—around 20%—compared to the stable state of the flock using optic distance. In both instances, the visual attraction function behaves as intended on a flock level, but the way that it interacts on a particle level is completely different. The difference in local interaction is rarely observed in traditional SPP models, mainly because most models have a region of interaction for the rules. This region of interaction restricts the model to only local information. When the model reacts only to local distance information, an implementation of attraction being proportional or inversely proportional (optic size) to distance averages at the same center of mass of the region for both cases. This differs from a visual implementation, where there is no single region. In the visual case, the difference in this relationship results in the center of mass changing from a local minimum to the true center of mass. This effect is clearly seen on the phase diagram of both implementations of the visual attraction rule (Fig. 2.9)

Note that for the exact same parameters, visual attraction based on optic size strug-

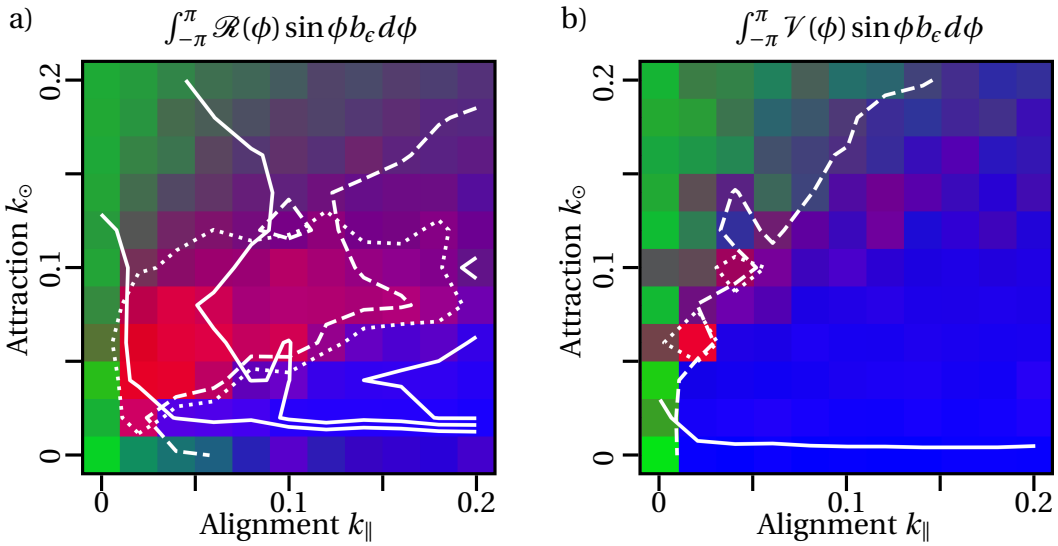

Figure 2.9.: Phase diagram N=50, different visual attractions
Phase diagram N=50, for different visual attractions and $\epsilon = 1$. a) Optic distance b) Optic size.

gles to reproduce milling. Further analysis shows that optic size cannot reproduce milling. The literature supports that milling is a phase that relies on a delicate balance of attractive forces (Strömbom 2011; Delcourt et al. 2016; Zienkiewicz et al. 2018). Furthermore, Papadopoulou et al. (2023b) and Costanzo et al. (2018) show that milling in fish locates the closest neighbors to the front of the individual. Fig. 2.10 shows what would be the average optic location seen by an individual during a stable milling phase. If you also consider that optic size tends to subdivide the flock into self-attracting smaller groups, the absence of milling when using optic size aligns with real-world data and statistical analysis.

One important point is that several traditional models use an attraction rule as an inversely proportional function of distance, which, from a biological point of view, is not feasible. However, the assumptions that were used to formulate the model resulted in the same behavior—a behavior that could not be replicated using biologically plausible assumptions. The model that was developed during this research allows for better exploration of collective motion without requiring a complete redefinition of traditional models. I managed to introduce the concept of what is possible to find in nature, what is possible to measure in nature, and what can be used in nature within a realm of mass-less particles and reconcile them. All of this was done without requiring complex spatial derivatives, occlusion-based filtering, or diminishing the impact of visual information on the behaviors of the animals.

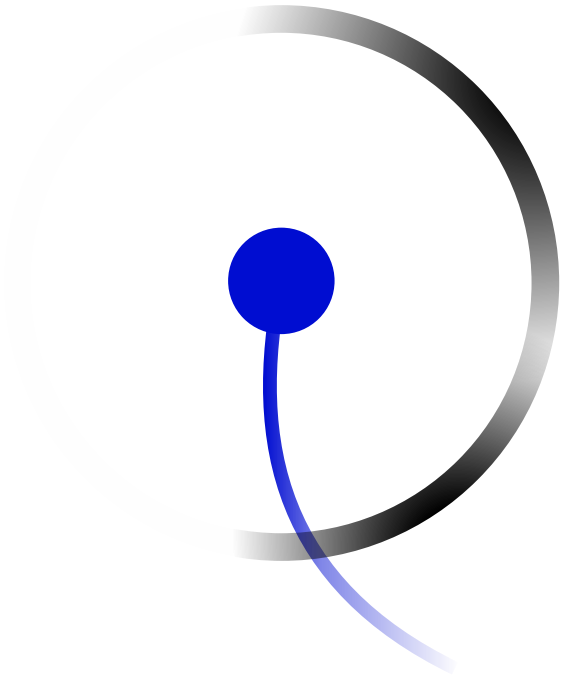

Figure 2.10.: Optic location seen by an individual during a stable milling phase.

3. Robotic self-propelled particles.

Contents

3.1. Abstract . . . 89
3.2. Introduction . . . 89
3.2.1. Related Work . . . 89
3.2.2. Contribution . . . 91
3.3. Methods . . . 91
3.3.1. Robotic Implementation Challenges . . . 91
3.3.2. Robotic Setup . . . 92
3.3.3. Anchoring the collective behavior . . . 95
3.3.4. Robotic Visual Flocking Model . . . 96
3.3.5. Robots-in-the-loop computer architecture . . . 98
3.3.6. Metrics . . . 100
3.4. Results . . . 101
3.4.1. Simulation of Supplemented Visual Model . . . 101
3.4.2. Phase Diagrams of the Supplemented Model in Simulation . . . 102
3.4.3. Comparison with the phase diagram of the Spheros' robots-in-the-loop . . . 103
3.5. Discussion . . . 103
3.6. Conclusion . . . 106
3.7. Supplementary Information . . . 109
3.7.1. Spheros' dynamical response and time delay τ estimation . . . 109
3.7.2. Direct comparison between Supplemented visual model and robot-in-the-loop Experiments . . . 110

We are standing on the shoulders of giants- Isaac Newton.

Chapter 2 described a collective motion model from a visual perspective. The visual model was able to reproduce the primary emergent phases of collective motion—Swarming, Schooling, and Milling. Furthermore, it was able to reproduce them

with the same formulation as traditional models. In other words, the proposed visual model had the same formulation, the same rules, and could be easily compared to traditional SPP models. However, the real world is not composed of mass-less particles moving in a vacuum (at least not on the scale we interact with every day).

Next, the implementation of the visual model on spherical robots is presented. The implementation of the visual model is done through a robots-in-the-loop approach. The model is supplemented to include some of the dynamics that a robotic implementation entails. With this addition, the model and the real-world experiments align, allowing a real-world reproduction of SPP models within the presented robots-in-the-loop approach. The phases reproduced and their distribution in the state-space closely match both simulation and real-world experiments.

Visual collective behaviors on spherical robots

Castro, D., Eloy, C., & Ruffier, F. (2025). "Visual collective behaviors on spherical robots." *Biomimetics & Bioinspircation*, 20(2), 026006.

3.1. Abstract

The implementation of collective motion, traditionally, disregard the limited sensing capabilities of an individual, to instead assuming an omniscient perception of the environment. This study implements a visual flocking model in a “robot-in-the-loop” approach to reproduce these behaviors with a flock composed of 10 independent spherical robots. The model achieves robotic collective motion by only using panoramic visual information of each robot, such as retinal position, optical size and optic flow of the neighboring robots. We introduce a virtual anchor to confine the collective robotic movements so to avoid wall interactions. For the first time, a simple visual robot-in-the-loop approach succeed in reproducing several collective motion phases, in particular, swarming, and milling. Another milestone achieved with by this model is bridging the gap between simulation and physical experiments by demonstrating nearly identical behaviors in both environments with the same visual model. To conclude, we show that our minimal visual collective motion model is sufficient to recreate most collective behaviors on a robot-in-the-loop system that be implemented using several individuals, behaves as numerical simulations predict and is easily comparable to traditional models.

3.2. Introduction

3.2.1. Related Work

Biological collective motion has inspired extensive efforts to understand, replicate, and utilize it in artificial systems. Nearly four decades ago, C. W. Reynolds (1987)

proposed that collective motion could emerge from simple, individual-level rules: (i) **attraction**, the tendency to stay close to others; (ii) **alignment**, the inclination to move in the same direction as neighbors; and (iii) **avoidance**, the need to prevent collisions. Vicsek et al. (1995) introduced a statistical model where individuals align their velocities with neighbors within a given proximity, incorporating stochastic noise to reflect individual variation. Couzin et al. (2002) expanded these ideas by integrating spatial interaction zones, with avoidance in the innermost region, alignment in the intermediate zone, and attraction in the outermost zone.

These **A**ttraction, **A**voidance, **A**lignment rules ("**3A**") effectively reproduce key collective behavioral phases observed in nature. We define these phases as: **swarming**[1] (non-polarized cohesive motion), **schooling** (highly polarized motion, also known as flocking), and **milling** (circular motion). Models based on these principles are commonly referred to as "3A models" **aoki1982simulation**; Mishra et al. 2010; Vicsek et al. 2012; Solon et al. 2015.

Many "3A models" rely on artificial methods to enforce rules, such as inter-individual distances **ashraf2017simple**; Chaté et al. 2008; Bialek et al. 2012; Wen et al. 2012, global velocity measurements Cavagna et al. 2015; H.-T. Zhang et al. 2010, or absolute communication networks Shang et al. 2014; Virágh et al. 2014. However, such approaches often diverge from the mechanisms used by animals, prompting researchers to explore more biologically plausible ("bio-plausible") models Lazarus 1979; Ginelli et al. 2010; Nagy et al. 2010; Bode et al. 2011a. Vision-based mechanisms have proven particularly effective, as vision is the primary sensory modality for many species Strandburg-Peshkin et al. 2013; Davidson et al. 2021; Soria et al. 2021; Lafoux et al. 2023.

Vision-based collective motion models have previously been tested on robotic platforms Schilling et al. 2022; Mezey et al. 2024. Moreover, some authors have previously used only vision to successfully model collective motion Collignon et al. 2016; Bastien et al. 2020; Krongauz et al. 2023; Krongauz et al. 2024. Some implementations treated vision as a selective mechanism D. J. G. Pearce et al. 2014; Soria et al. 2019 or do not fully adhere to the fundamental principles of the classical 3A model by not using its structure of "alignment", "attraction" or "avoidance" rules Bastien et al. 2020; Ito et al. 2024. Recently, Castro *et al.* (2024) Castro et al. 2024b proposed a model that relies solely on visual information while maintaining the structure of 3A models. Their model replicates collective behaviors using "early vision" cues : (i) angular position, (ii) angular size in the field of view as well as (iii) **optic flow**, the visual perception of the apparent angular motion of the environment caused by relative movement. Optic flow is a key navigational tool in nature, used by insects during flight Franceschini et al. 2007; Portelli et al. 2010 and by fish and birds Serres et al. 2019; Bonnen 2023. Robotic implementations of optic flow have also demonstrated significant potential Ruffier et al. 2005; Floreano et al. 2013; Expert et al. 2015; Bergantin et al. 2021; De Croon et al. 2022.

[1] Note, this "swarming" phase is a different concept to *swarm robotics*. Here, the **swarming** phase is an emerging behavior that occurs when the group moves in a non-organized motion.

3.2.2. Contribution

This study proposes a **robot-in-the-loop** approach to collective motion that relies exclusively on these “early vision" cues using educational spherical robots. Unlike previous models, this approach avoids reliance on global information or complex communication networks. Instead, it introduces a **visual anchor** to address environmental boundaries, eliminating the need for explicit wall modeling. Through numerical simulations and experiments involving 10 independently controlled robots, the method successfully reproduces diverse collective motion behaviors.

The article is organized as follows: Section 3.3.1 provides an overview of the challenges of a robotic implementation of 3A-models, with the example of the visual-based collective motion model proposed by Castro et al. (2024b) (Fig.3.1). Section 3.3.2 describes the robotic platform and the model modifications for implementing on the 10 independently controlled robots including the introduction of the visual anchor concept. Section 3.3.4 detail the implementation of a feature matching optic flow algorithm and visual collective motion rules. Section 3.3.5 describes the supplementation on original model and summarizes robotic setup, while Section 3.4 presents the results from simulations and robot-in-the-loop experiments. Finally, Section 3.5 concludes with a discussion of the results and directions for future research.

3.3. Methods

3.3.1. Robotic Implementation Challenges

Flocking models based on individual rules of attraction, avoidance, and alignment – commonly referred to as 3A-models – pose several challenges when implemented on robotic platforms. Two significant challenges arise due to differences between theoretical models and real-world implementations.

The first challenge is related to the abstraction of individual agents. Most 3A-models assume that individuals are self-propelled particles moving in a plane. However, many robots either deviate from the particle abstraction or lack natural constraints to planar motion. This raises the question: *Which robot kinematics should be chosen?* The chosen robotic platform must strike a balance between adhering to model assumptions, ensuring adequate sensing capabilities, accommodating computational constraints, and maintaining a suitable size, among other factors.

The second challenge involves transitioning from numerical simulations to physical experiments. Unlike numerical environments, experimental setups are spatially constrained. As a result, collective behaviors must be confined within the experimental area. Without addressing these constraints, implementing flocking behaviors on robots would merely involve a direct adaptation of an established model, such as the one proposed by Castro et al. (2024b), to the chosen robotic platform (illustrated in Fig. 3.1).

The following sections present our solutions to these challenges.

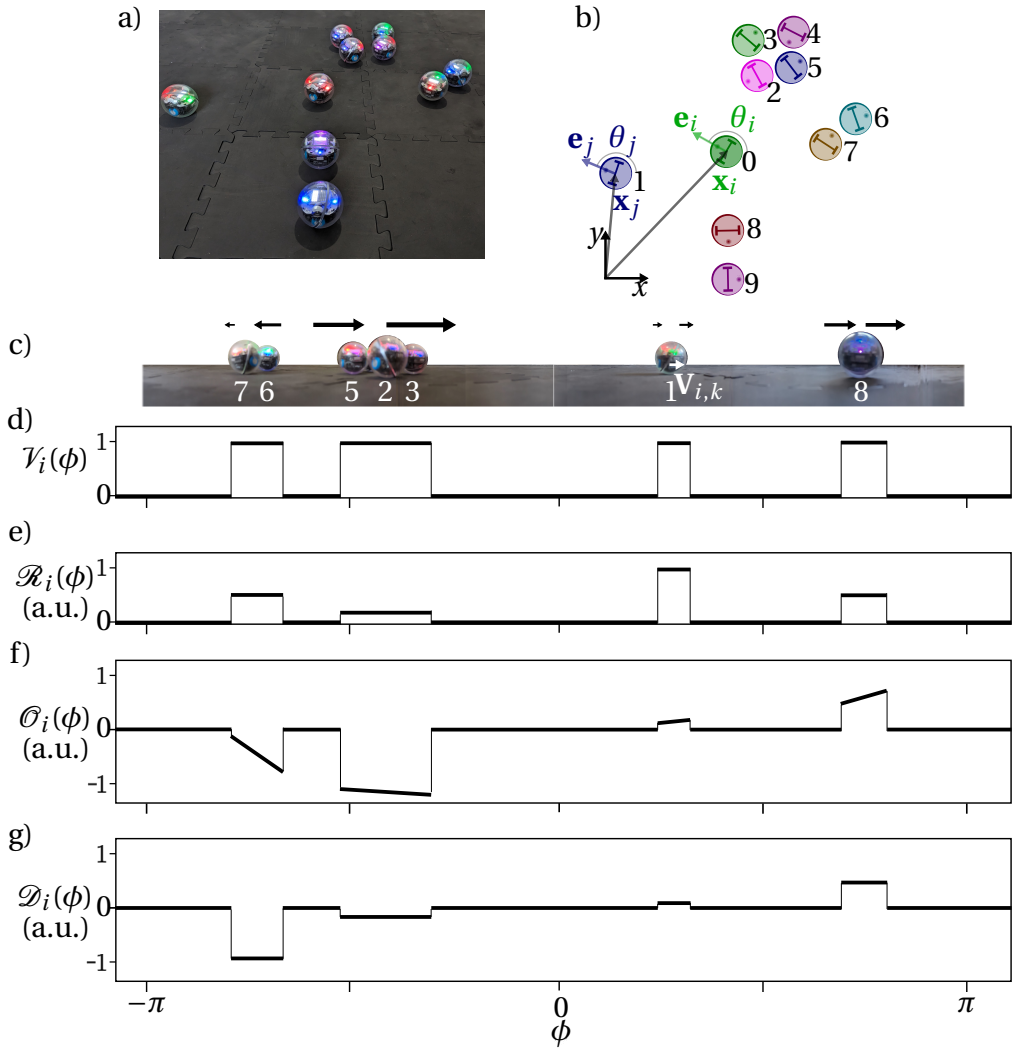

Figure 3.1.: Modeling collective behaviors from optic flow and retinal cues Castro et al. 2024b: a) View of the flock positions. b) Representation of flock positions after recognition at time step t. Each i-particle is described by its XY-coordinates and heading unit vector $\mathbf{e}_i$, with heading angle θ_i. c) Point-of-view of particle *0*, showing the relative retinal positions of each particle. The arrows represent the perceived optic flow, with the white arrow indicating the apparent velocity ($\mathbf{V}_{i,k}$). Only the azimuthal component of $\mathbf{V}_{i,k}$ is shown for retinal object k as perceived by $i = 0$. d) $\mathcal{V}(\phi)$: Binary panorama representation of the field of view of particle *0*. e) Apparent distance $\mathscr{R}_i(\phi)$ for particle *0*. f) Optic flow $\mathscr{O}_i(\phi)$ perceived by particle *1*. g) Optic flow divergence $\mathscr{D}_i(\phi)$ perceived by particle *1*.

3.3.2. Robotic Setup

This section describes the robotic setup for implementing a robot-in-the-loop vi-

sual flocking model. We propose to use educational spherical robots equipped with wireless control capabilities, while the robot's sensing capabilities are provided by an external infrastructure. The external infrastructure reconstructed each binary panorama and processed the model described in Section 3.3.1, controlling the locomotion of each robot independently.

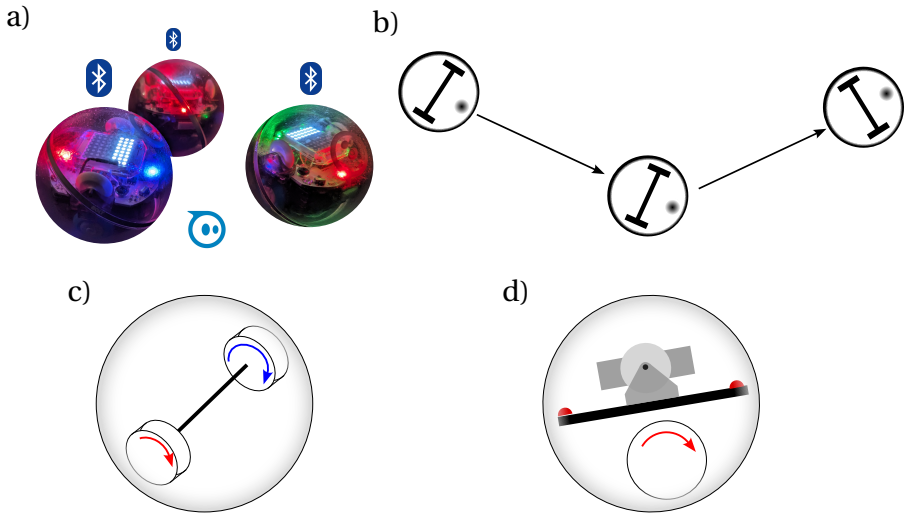

Figure 3.2.: Spherical robotic platform. a) Sphero Bolt ® wireless educational robots. b) Example of the robots' movement, c) Heading control by differential rotation on the two opposite inner wheels, d) Induced nose-up pitch during forward displacement.

This robot-in-the-loop solution leverages an existing educational robotic platform, the Sphero Bolt® (Fig. 3.2a). These compact, spherical robots can be controlled in real-time via Bluetooth®using an educational application, which allows users to set a heading, desired velocity, and LED states, as well as read onboard sensor data. Furthermore, several open-source projects extend the functionality of these robots by integrating them with programming environments such as ROS, Python, and/or MATLAB®.

Each robot is equipped with a suite of sensors, including an Inertial Measurement Unit (IMU) composed of a 3-axis accelerometer, gyroscope, and magnetometer, as well as a light sensor. Additionally, the robot features four infrared (IR) proximity sensors positioned on its sides, two addressable RGB LEDs (located at the front and rear), and an 8×8 RGB LED matrix on the top.

Locomotion is facilitated by a self-balancing two-wheel enclosed system, with two passive wheels (not shown in Fig.3.2) maintaining continuous contact with the spherical enclosure. This dual-motor configuration enables linear movement, while stationary turning is achieved via differential motion (Fig.3.2 b)-c)). The locomotion process induces a pitch displacement dependent on the robot's speed and the friction coefficient between the spherical enclosure and the surface (Fig.3.2 d)).

This pitch displacement is partially mitigated by the robot's dual-MCU architecture: one microcontroller unit (MCU) manages the communication protocol and high-level

commands, while the other is dedicated to motor control. The latter ensures precise implementation of the desired heading and speed while handling low-level sensor readings.

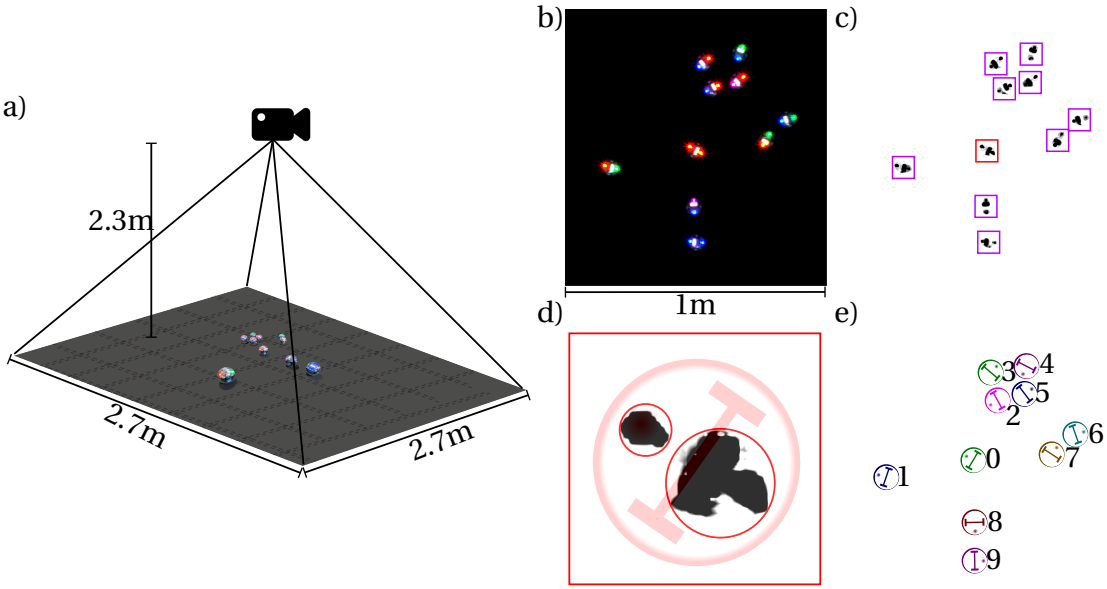

Figure 3.3.: Experimental Setup and external infrastructure. a) Photograph of the experimental setup. b) Sample image captured by the camera (Daheng Imaging MER2-230-168U3C camera equipped with a 4 mm C-mount lens). c) Whole-robot body image segmentation: Segmentation of the robot's body based on the captured image. d) Recognition of individual light sources and heading: Identification of the front and rear LEDs allows calculation of the robot's heading vector. e) Robot identification and virtualization: Assignment of a unique identifier to each robot and visualization of their positions and orientations.

To address the absence of onboard cameras and the inability to integrate them into the robots, an external infrastructure made of a bird's-eye camera was positioned above the experimental area. The camera used is an industrial RGB global shutter model, the *Daheng Imaging MER2-230-168U3C*, with a frame rate of 168 frames per second (FPS), a resolution of 1920×1200 pixels, and compatibility with interchangeable C-mount lenses (Table 3.1). A 4 mm, 1/8", F2.0–12 non-distorting lens was selected, enabling the camera to be mounted 2.3 m above the robots while covering a 2.7×2.7 m working area—approximately 38 times the length of a single robot (Fig. 3.3 a)).

Each robot is assigned a unique color pattern, enabling the identification of its front and back. To maintain visual connectivity between these sections, the back half of the LED matrix is set to emit 50% bright white light. This configuration supports 16 distinct combinations for robot identification, with potential for further expansion using additional colors.

A bird's-eye camera captures a 1920×1200 RGB image, enabling the isolation of each light source on the robots and reduces light bleeding when several robots are in close

Resolution	1920x1200
Frame rate	168 FPS
Shutter time	25 ms
Pixel Bit Depth	10 bit
Digital gain	5 dB
Pixel Data Format	Bayer RG10
Lens f-stop	11
Interface	Python 3.10
Non-distorting lens	4mm (focal length), 1/8" (optical format) , F2.0-12 (aperture)

Table 3.1.: *Daheng Imaging MER2-230-168U3C* global shutter Camera Settings.

proximity (Fig. 3.3 b)). Whole-robot segmentation is performed on this image, treating the three light sources on each robot as a single entity. This allows for the calculation of the robot's center coordinates (X,Y), which approximate its geometric center. By knowing the radius of the robots in pixels, N cropped images are extracted from these coordinates, each isolating an individual robot.

The cropped images are further segmented to recognize the front and back LEDs as well as the LED matrix. The larger area created by the two light sources facilitates accurate recognition of the robot's front and back (Fig. 3.3 d)). Additionally, the unique color patterns from this segmentation are used to assign a distinct identifier to each robot.

Finally, the robot's heading is calculated by tracing the vector formed between the center positions of the front and back LEDs. This computer vision processing yields the robot's identification, X,Y-coordinates, and heading (Fig. 3.3 e)).

The next phase, following the identification and localization of each robot, involves reconstructing the binary panorama for each robot.

3.3.3. Anchoring the collective behavior

In the binary panorama reconstruction, we introduce a novel component to the original model Castro et al. 2024b: a visual anchor. The visual anchor acts as an attractive force directed toward a designated location, which, in this case, is the center of the arena. The farther an individual is from this location, the stronger the anchor's influence, causing the individual to turn and face the anchor. The anchor is described by Eq. (3.1), where $\mathscr{R}_i$ represents the apparent distance function over the binary panorama observed by particle i, and a denotes the particle's radius.

$$\omega_c = \int_{-\pi}^{\pi} \frac{\mathscr{R}_i^2(\phi)\sin(\phi)}{a} d\phi, \tag{3.1}$$

The inclusion of an anchor bounds the flock to a place and prevent infinite expansion. The anchor is an augmented reality (AR) solution that enables each robot to see the desired center point (Fig. 3.4 a)). This anchor is omnisciently visible, regardless of any

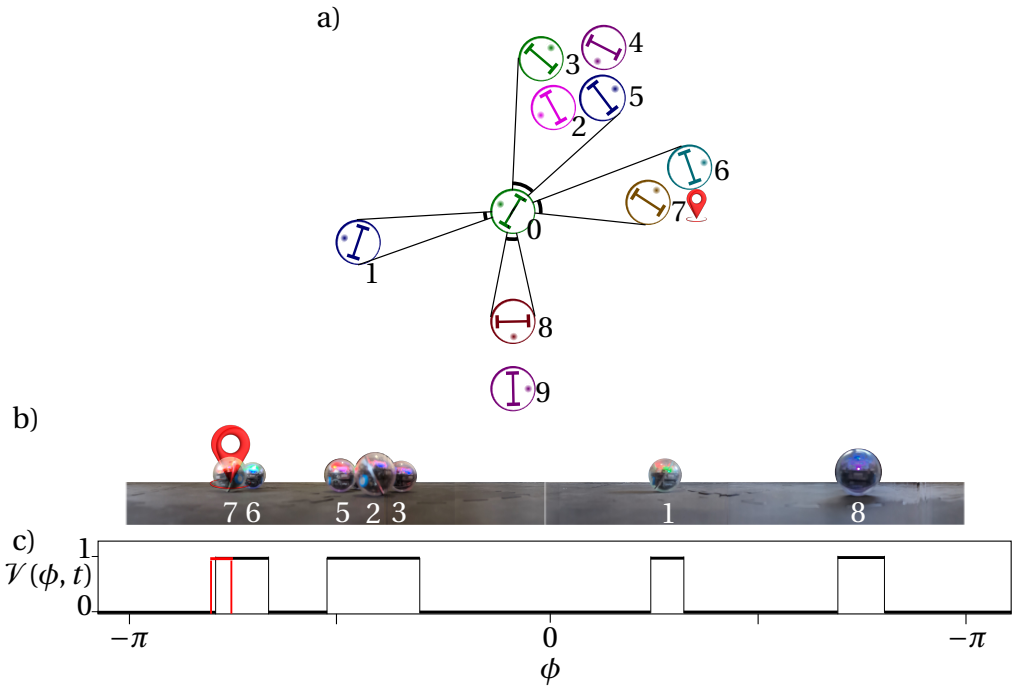

Figure 3.4.: Binary Panorama Reconstruction. a) 2D virtual representation of the binary panorama: The red goal indicates the desired location of the flock. b) Mixed-reality perspective from particle 0: Displays the omniscient red goal as observed from the particle's point of view. c) Graph representation of all $\mathcal{V}(\phi, t)$: $\mathcal{V}(\phi, t)$ represents the binary panorama view of the field of vision for particle i, with the anchor displayed as a distinct channel (red). This graph can be efficiently stored as a tuple list, representing the rising and falling edges of the binary panorama.

obstacles between the line of sight and the anchor location (Fig. 3.1 b)). It is a new visual channel that uses the same binary representation as the visual surroundings (Fig. 3.1 c)). The binary panorama is represented in Python as an ordered tuple list of every retinal object's rising and falling edge, with the last position always being the goal tuple. This process is performed independently for each particle without sharing information between particles.

3.3.4. Robotic Visual Flocking Model

Avoidance effect ($\mathscr{A}$) is here introduced in the attraction term ($\omega_{\odot}$). Our previous study Castro et al. 2024b did not include any avoidance because most collective motion models are based on the assumption that all particles have some extra room on the z-axis, allowing them to avoid collisions with additional depth. This assumption is primarily based on how fish flocking data is gathered, where the fish can swim above or below each other in a few centimeters of water Sara Brin Rosenthal et al. 2015. However, for any robotic implementation, obstacle or collision avoidance is of critical importance. Fortunately, avoidance and attraction have similar characteristics, being

proportional instantaneous responses to the optical size and location of a retinal object. The key difference is in the direction of the response: while attraction is positive, avoidance is negative. Therefore, avoidance is implemented by providing a new piecewise constant function ($\mathscr{A}$), where attraction is inverted in situations where a particle should avoid a retinal object. This new function $\mathscr{A}$ depends on the width of the retinal object (optical size), specifically the difference between its rising and falling edges ($\Delta\phi$). Note that the apparent distance of a retinal object ($\mathscr{R}(\phi,t)$) inversely depends on this optical size $\Delta\phi$. We choose $\mathscr{A}(\Delta\phi) = -5$ for $\Delta\phi < \pi/10$, which is equivalent to the retinal object being closer than three particle radii ($\mathscr{R} < 3a$). In all other cases, $\mathscr{A}(\Delta\phi) = 1$. The blind spot remains unchanged, giving more weight to the retinal object in front of the particle, regardless of this new function.

Alignment ($\omega_\parallel$) remains unchanged with respect to previous study Castro et al. 2024b, considering that the divergence/convergence of the optic flow already accounts for retinal objects approaching the focal object.

The optic flow for each retinal object can be decomposed in two components: the azimuthal in its center ($\overline{\mathscr{O}}$) and its radial divergence ($\mathscr{D}$). The former represents the apparent displacement of retinal features, while the latter represents changes in the size of those features. Divergence is a decrease in size, and convergence is an increase. With a 1D representation, these features are the rising and falling edges of the retinal object, akin to a feature-matching optic flow calculation.

Calculating optic flow as a feature-matching operation heavily relies on accurately identifying the same features across consecutive time steps. The model accomplishes this by finding the gradient of the Euclidean distance of the retinal object position and size between consecutive time steps. Regardless of flock size, there are three possible cases. If, in both time steps, there are the same number of retinal objects, the minimum of the gradient will be continuous and located on the main diagonal of a square matrix. When the number of retinal objects differs, it indicates either a separation or merge of retinal objects. In this case, there is no match, and the retinal objects are ignored in the calculation.

The apparent velocity vector $\mathbf{V}_{ik}$ in polar coordinates is $\mathbf{V}_{ik} = [-\mathscr{D}_i, \overline{\mathscr{O}}_i] \cdot \mathscr{R}_i / U$ with the heading vector $\mathbf{e}_i = [\cos\phi, -\sin\phi]$. The supplemented visual model equations for the collective behavior are presented in Eqs. (3.2)-(3.7). Note that the noise k_η and the time delay τ are set to 0 for robot-in-the-loop experiments, as they are parameter modeling real-world noise and uncertainty.

$$\dot{\mathbf{x}}_i(t) = U\mathbf{e}_i(t), \tag{3.2}$$

$$\dot{\theta}_i(t) = k_\odot\,\omega_\odot(t) + k_\parallel\,\omega_\parallel(t) + k_c\,\omega_c(t) + k_\eta\,\eta(t), \tag{3.3}$$

$$\omega_\odot(t) = \left\langle \int_{-\pi}^{\pi} \mathscr{R}_i^2(\phi, t-\tau)\mathscr{A}(\mathscr{R}_i(t-\tau))\,b_\epsilon(\phi, t-\tau)\sin\big(\phi, t-\tau\big)\,d\phi \right\rangle, \tag{3.4}$$

$$\omega_\parallel(t) = \left\langle \int_{-\pi}^{\pi} \frac{\mathbf{e}_i(t-\tau)\times\mathbf{V}_{ik}(t-\tau)}{U\mathscr{R}_i(\phi, t-\tau)}\cdot\hat{\mathbf{z}}\,b_\epsilon(\phi, t-\tau)\,d\phi \right\rangle, \tag{3.5}$$

$$\omega_c(t) = \int_{-\pi}^{\pi} \frac{\mathscr{R}_i^2(\phi, t-\tau)\sin\big(\phi, t-\tau\big)}{a}\,d\phi, \tag{3.6}$$

$$\frac{\mathbf{e}_i(t)\times\mathbf{V}_{ik}(t)}{\mathscr{R}_i(t)}\cdot\hat{\mathbf{z}} = -\mathscr{D}_i(\phi, t)\sin\big(\phi, t\big) + \overline{\mathscr{O}}_i(\phi, t)\cos\big(\phi, t\big). \tag{3.7}$$

Where $\mathbf{x}_i(t)$ is the XY-coordinate of particle i at time t, $\mathbf{e}_i(t)$ is the heading vector of particle i at time t, $k_\odot$, $k_\parallel$, k_c, k_η are the gains of the attraction, parallel, anchor and noise respectively while $\omega_\odot(t)$, $\omega_\parallel(t)$, $\omega_c(t)$, η are the attraction, parallel, anchor and noise terms at time$t$. $\hat{\mathbf{z}}$ is the unitary vector that is orthogonal to the ($\hat{\mathbf{x}}$, $\hat{\mathbf{y}}$)-plane of movement of the robots. ϕ is the retinal angle in the field of view, is considered from the point of view of each particle and counter-clockwise is considered positive. $\mathscr{R}_i(\phi, t)$ is the apparent distance function at a retinal angle ϕ at time t for individual i. $b_\epsilon(\phi, t)$ is the blindspot function $(1+\epsilon\cos\phi)$. $\mathbf{V}_{ik}(t)$ is the apparent velocity that an individual i perceives of a retinal object k. Finally τ is the delay between perception and action.

The robotic platform introduces several critical processing factors – Bluetooth, image processing, internal Spheros control command, ... – that affect the overall dynamical response of the robots and its collective behavior as a whole (see Sec. 3.7.1). These extra factors results in a time delay τ with a median of 600ms. The overall functional diagram of the collective motion model is presented in Fig. 3.5. This model is executed independently and simultaneously for all robots; it does not involve communication between robots, sharing information, and only uses visual data. The same block could be utilized on a robotic platform using LiDAR, onboard cameras, touch sensors, ultrasound, or any other sensor that provides a binary visual panorama as output. All the information used by this model is self-referred and self-contained; there is no global reference frame, coordination, or knowledge of the flock or the environment.

3.3.5. Robots-in-the-loop computer architecture

We propose a robots-in-the-loop flocking framework that consists of a set of N Sphero® Bolt robots, which lack visual sensors and are controlled via Bluetooth®. Their positions and headings are determined by processing images captured by an industrial 1920x1200 RGB, 164 FPS overhead camera. From these positions, a binary panorama is emulated. This binary panorama serves as the input to the collective

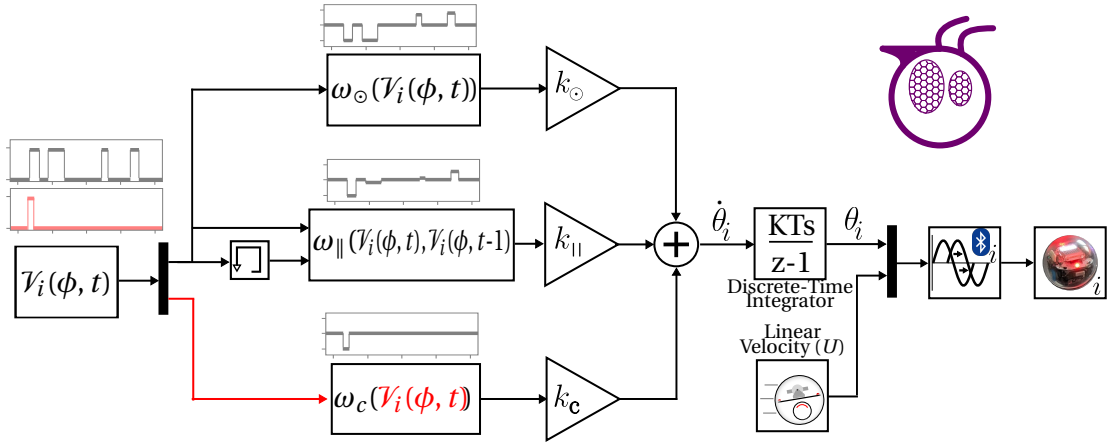

Figure 3.5.: Functional diagram of the collective motion algorithm with robots-in-the-loop.

motion model, as shown in Fig. 3.6.

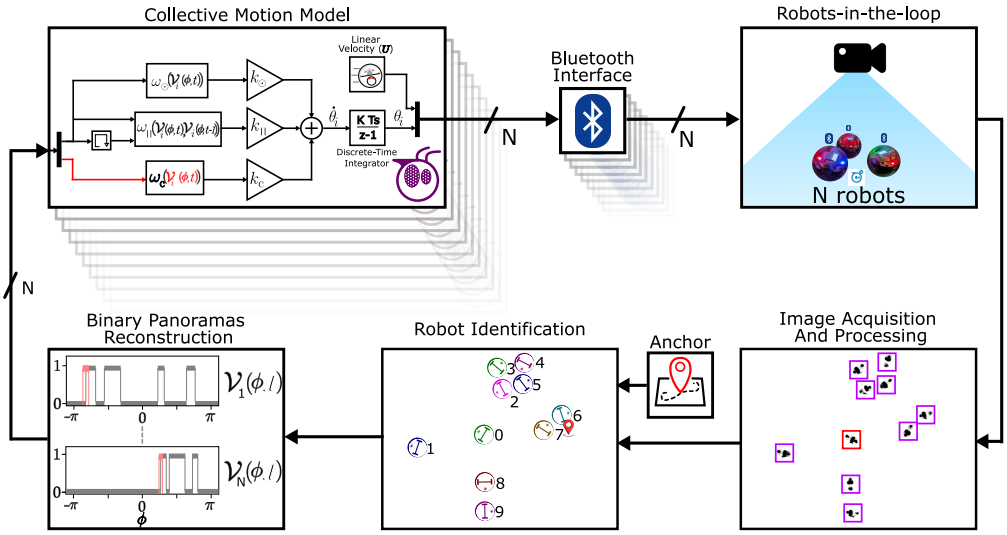

Figure 3.6.: General Block Diagram. The N flocking robots are observed using a 1920x1200 px, 164 fps camera. The captured images are processed to determine the XY-coordinates, heading, and colors of all N robots. Based on these coordinates, each robot is uniquely identified. The XY-coordinates of the Anchor can be set at any point within the image to assist in robot identification. These coordinates are further processed to reconstruct the binary panorama for each spherical robot. Each binary panorama serves as input to an independent collective motion model, which processes the visual panoramic data. These models operate in parallel, and their outputs, in terms of heading, are transmitted to each robot via a Bluetooth interface.

There are two main Python programs to handle the implementation of the blocks, as shown in Fig. 3.6. The first one (*administrator.py*) coordinates the camera communication, data extraction, image processing, object identification, and binary panorama reconstruction (Fig. 3.3). The binary panorama reconstruction is transferred to N independent routines named *PyRobot.py* (the second program). Each routine processes the binary panorama of a single robot to calculate its output from the collective motion model and send the desired heading and speed via Bluetooth to that same robot (Fig. 3.5). Each robot receives the inputs (heading and velocity) and moves accordingly. Each *PyRobot.py* is a routine that can be executed either locally or remotely from the administrator. This setup allows a framework only bounded in the flock size to the maximum number that can fit within the camera's recognition area and hardware-constrained Bluetooth interface capabilities of the computer. For *Windows*® machines, the theoretical maximum is seven devices per machine; in *Linux*, it is seven per interface. However, in practice, it is recommended to have a maximum of five on the same machine managing the camera and six per interface on independent machines.

3.3.6. Metrics

To quantitatively distinguish between different dynamical phases, we introduce three global order metrics: polarization (P), milling (M), and opacity (O) Calovi et al. 2014; Filella et al. 2018. These metrics are defined as follows:

$$P(t) = \left\| \overline{\mathbf{e}_i(t)} \right\|, \tag{3.8}$$

$$M(t) = \left\| \overline{\mathbf{y}_i(t) \times \mathbf{e}_i(t)} \right\|, \tag{3.9}$$

$$O(t) = \frac{1}{2\pi} \overline{\int_{-\pi}^{\pi} \mathcal{V}_i(\phi, t)\, d\phi}, \tag{3.10}$$

where the overbar denotes an average over all individuals, and the unit vector $\mathbf{y}_i(t) = \frac{\mathbf{x}_i(t) - \overline{\mathbf{x}_i}(t)}{\|\mathbf{x}_i(t) - \overline{\mathbf{x}_i}(t)\|}$ points towards particle i from the center of mass. All three metrics range within the interval $[0,1]$. The polarization $P(t)$ measures alignment: $P(t) = 0$ corresponds to particles pointing in all directions, while $P(t) = 1$ represents a perfectly aligned school. The milling $M(t)$ quantifies the normalized angular momentum: $M(t) = 0$ corresponds to a straight-line formation, and $M(t) = 1$ represents perfect milling. The opacity $O(t)$ measures the "occupancy" of the visual fields: $O(t) = 0$ indicates no object in the visual field, and $O(t) = 1$ signifies that the entire visual field is obscured. To asses the performance of the phase, we quantify the number of collisions while a simulation or experiment were running. For an experiment, we counted how many frames presented a collision and how many robots were involved. For a simulation we counted collision by defining one as any time the distance between individuals was less than $2a$.

3.4. Results

3.4.1. Simulation of Supplemented Visual Model

To make the problem dimensionless, we chose $a = 1$ and $U = 1$. The blind angle parameter and noise strength were set to values that enable the reproduction of most phases: $\epsilon = 1$, $k_\eta = 0.01$. With this approach, three dimensionless parameters remain: the strengths of anchor attraction and alignment, k_c, $k_\odot$, and $k_\parallel$. The anchor parameter was chosen to confine robotic experimentation, set practically to the minimum value required to prevent the robots from reaching the arena's border when no other law is in effect, $k_c = 0.005$.

The original visual model has been supplemented with three significant modifications, influencing the phase distribution and metrics in the parameter space $(k_\odot, k_\parallel)$:

- inclusion of the avoidance rule ($\mathscr{A}(\phi)$) (Fig. B.5),
- addition of an anchor law (ω_c) (Fig. B.6),
- introduction of a time delay (τ) (simulation only) to reflect the delay between the desired heading and a perceived change in heading when operating the Spheros robot (Fig.3.7).

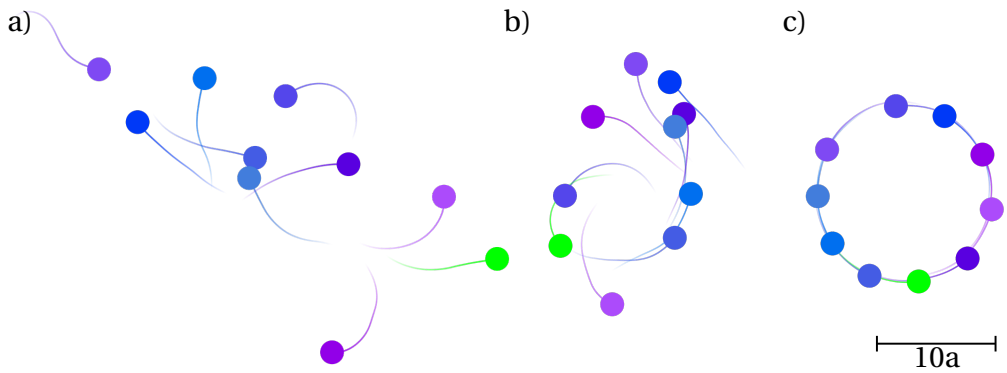

Figure 3.7.: Illustration of the phases observed in simulation for the supplemented visual model with the addition of avoidance, anchor, and time delay as defined in Eqs. 3.2, with $\tau = 2.4[a/U]$. Three phases are observed when varying the other two parameters: a) Swarming ($k_\odot = 0.25$, $k_\parallel = 0.04$) (see SVideo 1); b) Bistable ($k_\odot = 0.2$, $k_\parallel = 0.25$) (see SVideo 2); c) Milling ($k_\odot = 0.07$, $k_\parallel = 0.16$) (see SVideo 3).

We compared the original model with each successive addition by exploring the parameter space $(k_\odot, k_\parallel) \in [0, 0.3] \times [0, 0.3]$ for a flock of 10 individuals. For each parameter set, we ran 10 simulations over an extended duration ($\Delta t = 5000$). The mean values of P, M, and O were determined by averaging over the final 1000 time

units to minimize the influence of transients. The simulation results are summarized in phase diagrams (Fig. 3.8).

We classified the collective phases using the polarization P and milling M parameters:

Schooling when $P > 0.5$ and $M < 0.5$, Milling when $P < 0.5$ and $M > 0.5$, Swarming when $P < 0.5$ and $M < 0.5$.

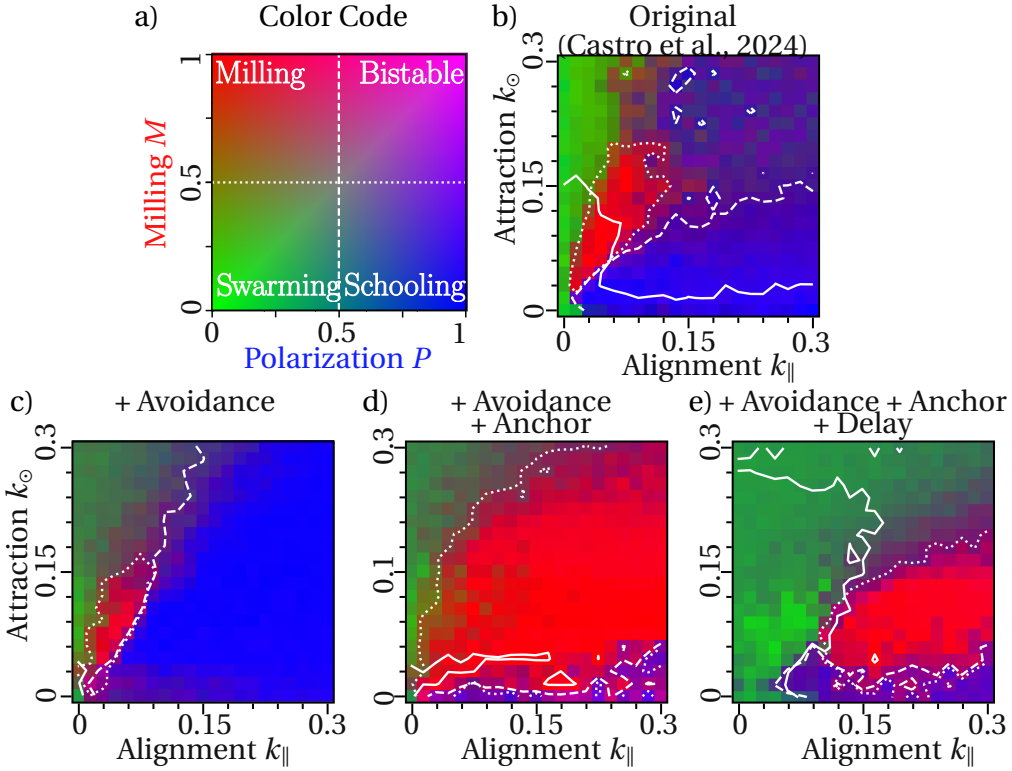

Figure 3.8.: Simulation Phase Diagrams for 10 individuals. The dotted line indicates $P = 0.5$, the dashed line indicates $M = 0.5$, and the solid line indicates $O = 0.3$. a) Color code, b) Original visual model Castro et al. 2024b, c) Supplemented visual model with avoidance term, $\mathscr{A}(\mathscr{R}(\phi, t))$, d) Supplemented model with avoidance and anchor term, $\omega_c(t)$, e) Supplemented visual model with avoidance, anchoring, and perceived time delay ($\tau = 2.4[a/U]$).

3.4.2. Phase Diagrams of the Supplemented Model in Simulation

We further compared the supplemented model and its robotic implementation by exploring the parameter space $(k_\odot, k_\parallel) \in [0, 0.3] \times [0, 0.3]$ for a flock of 10 independently controlled robots. For robotic experiments, dimensionless parameters were adjusted to match the radius in pixels and velocity in radius per second, resulting in $a = 1$, with other parameters remaining constant ($\epsilon = 1$, $k_\eta = 0$, and $k_c = 0.005$).

For each parameter set, five robotic experiments were conducted over an extended period (240 sec or $\Delta t \approx 8500$). The mean values of P, M, and O were calculated by averaging over the final 1000 time units to eliminate transient effects. The experimental results are summarized in phase diagrams (Fig.3.9 a). Examples of swarming, partial schooling in a bistable state, and milling are presented in Fig.3.9 b), Fig.3.9 c), and Fig.3.9 d), respectively. Note that the bistable state has been observed in fish experiments (Lafoux et al. 2024).

The phase diagram of the original model Castro et al. 2024b showcases the reproduction of the main three phases (swarming, schooling, and milling), with the area above $k_\odot > 0.15$ and $k_\parallel > 0.15$ being bistable (Fig. 3.8 a)). A bistable state is when the flock oscillates between milling and schooling (Fig. 3.8 b)). By adding repulsion to the model, this bistable phase is removed, but the area where milling appears shrinks to roughly half of the original while the final opacity of the flocks is reduced. Opacity is an inverse metric of flock density, and this reduction aligns with the expected effect of adding avoidance.

Fig. 3.8c shows the introduction of the anchor, which eliminates the schooling phase while the swarming phase area remains unchanged. This change is attributed to the turning effect generated by the anchor. When all particles are in a line, the front particles turn faster than those at the back, resulting in a turn that eventually stabilizes into a circle and a very stable milling pattern.

Adding a time delay ($\tau = 2.4[a/U]$) in the supplemented visual model (Fig. 3.8d) causes the swarming phase to expand to values above $k_\odot = 0.17$ or when $k_\parallel < 0.08$. For low attraction ($k_\odot < 0.03$), there is bistability, and the milling phase is reduced to an area for alignment above $k_\parallel = 0.1$ and medium attraction ($0.06 < k_\odot < 0.17$).

3.4.3. Comparison with the phase diagram of the Spheros' robots-in-the-loop

During the robotic implementation (see Fig. 3.9), the milling phase remains fairly stable, appearing for high alignment ($k_\parallel > 0.15$) across most attraction values ($k_\odot < 0.16$). The swarming phase is constant for low alignment ($k_\parallel < 0.06$) or high attraction ($k_\odot > 0.17$).

The avoidance law, subject to the same time delay, results in some collisions that resolve over time, leading to higher opacity values. Similarly, for $k_\odot < 0.01$, avoidance is too small, and the flock groups around the anchor without moving. This behavior is somewhat mitigated when there is enough alignment ($k_\parallel > 0.15$), but the flock remains more compact than in the simulations.

3.5. Discussion

In this article, we presented a robot-in-the-loop approach to a visual collective motion model. The robots used are practical to deploy and commercially available: the Sphero BOLT (Fig. 3.2). The Sphero BOLTs are sphere-like robots with a transparent shell

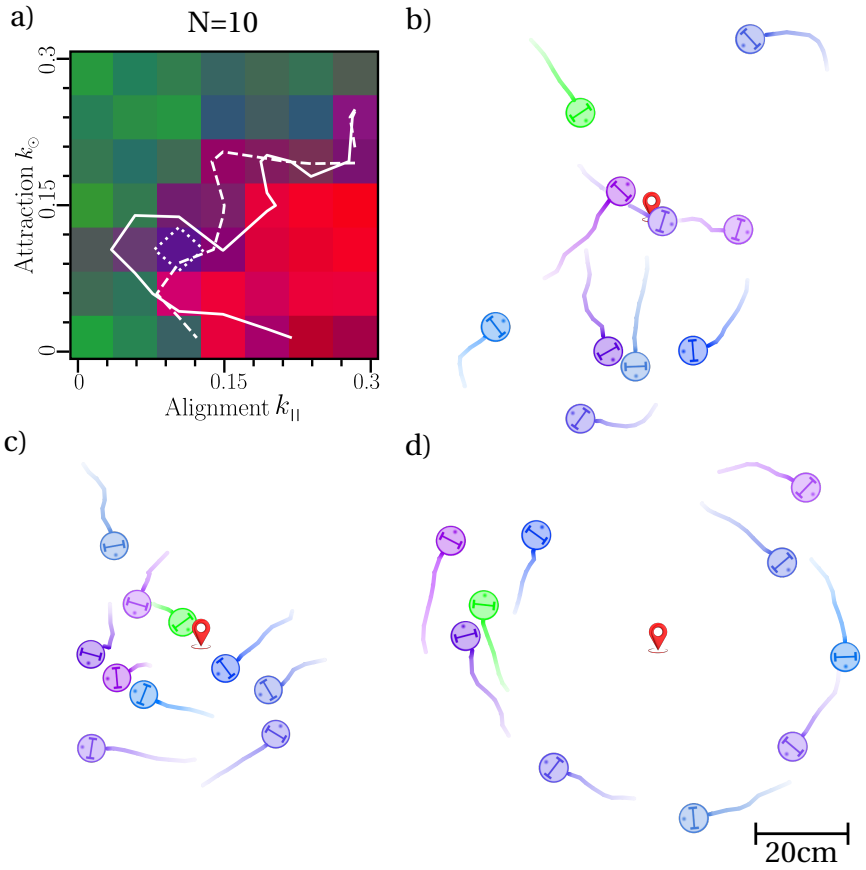

Figure 3.9.: Experimental results for a flock of 10 independently controlled robots (with $U = 4a/sec = 0.14[m/sec]$ (the radius of Spheros robot is $a = 3.5cm$). a) Phase diagram, the dotted line is $P = 0.5$, the dashed line is $M = 0.5$, and the solid line is $O = 0.7$. b) Swarming ($k_\odot = 0.25$, $k_\parallel = 0.04$) (see raw footage SVideo 4, robotic schematic SVideo 5 or a 3rd person view SVideo 6); c) Bistable ($k_\odot = 0.2$, $k_\parallel = 0.25$) (see raw footage SVideo 7, robotic schematic SVideo 8 or a 3rd person view SVideo 9); d) Milling ($k_\odot = 0.07$, $k_\parallel = 0.16$) (see raw footage SVideo 10, robotic schematic SVideo 11 or a 3rd person view SVideo 12).

that encloses all electronics, sensors, motors, batteries, and LEDs in a sealed design. However, due to this sealed design, integrating vision required an external solution. We opted for an external infrastructure made of overhead camera system to recognize the robots, enabling the reconstruction of each robot's visual field. Each robot was assigned a unique light pattern, visible through the transparent shell, to facilitate position and heading recognition via the overhead camera. (Fig. 3.3).

The noise effect The collective motion model presented by Castro et al. (2024b) and the supplemented version introduced in this article, incorporated noise as a component during the computation of angular velocity model output which is directly

used to actuate the robot. Introducing the noise on the actuation signal accounted for the noise that errors on the heading recognition might introduce, but did not accounted for recognition errors of the position. And, by extension, the noise that those errors might introduce on the robots' panorama reconstruction hence on the measurement of the "early vision" cues (apparent distance, optical position, and optic flow signals on Fig. 3.1). The effect of noise on these "early vision" cues can be particularly significant for the alignment term in the model. Alignment is calculated using optic flow and apparent distance (Eq. (3.5)), making it three times more sensitive to measurement noise than attraction that only uses apparent distance (Eq. (3.4)).

Fig. S1 illustrates the combined effect of these recognition errors in open loop, showing that the mean error is approximately 6%. Consequently, the attraction term inherently has a 6% error, while alignment error reaches around 15%. Despite this, the simulation and experimental robotic results align on a macro level as shown in Fig. 3.9. For comparison, the mean squared error (MSE) between the phase diagram images from robots-in-the-loop and supplemented visual model results is as low as 2.48% (for more details, see Fig. 3.10).

Effect of collisions We analyzed the number of collisions throughout the evolution of the flock during the examples of swarming, bistability, and milling. If a collision occurred during the robotic experiment or simulation, we counted how many individuals participated on the collision (as seen on Fig. B.7). In simulation, a collision is detected when the distance between individuals is less than or equal to $2a$. In the robotic experiment, a collision is detected when the recognized object becomes much larger than a single Sphero. These collisions are calculated throughout the simulation and the robotic experiment, and are presented as additional information on the state of the flock (Fig. B.8).

We found that, on average, in one video for each phase, the flock in simulation spent 14% of their time experiencing collisions. The robots exhibited a collision rate of 33% of the time (on average, three times more than the simulation). Interestingly, most simulation collisions were between two particles, and instances of collisions involving more than two particles were rare (only seen on Fig S9-b). In contrast, with the robots, most collisions caused a chain reaction, leading other particles to collide with the stopped particles (as seen on Fig S9-d). The most collisions occurred during the bistable state, where the simulation had a 21% collision rate compared to 36% for the robots. In both cases, the fewest collisions occurred during the milling phase. The simulation only showed a 6% collision rate, while the robots exhibited a 30% collision rate. The stability of the milling phase is also seen on the minimum inter-distance that, on simulation and on the robots, tends to stabilize around t=1200. The timeline of this phase can be seen on Fig S9-c for simulation and Fig S9-f for robots.

The anchor's effect One notable modification to the model was the introduction of an attractive force towards the center of the area of interest, referred to as the anchor (Eq. (3.1)). As shown in Fig. 3.8c, the main phase is schooling. However, when

the anchor is introduced (Fig. 3.8d), schooling disappears. We hypothesize that the anchor, acting as a force that causes individuals to turn towards the center when they are far from it, challenges the formation of schooling. Notably, schooling is typically observed during flock translation, such as animal migration, where the flock's center of mass moves along the flock's polarized direction.

If the model bounds a flock with a wall, a schooling flock would inevitably interact with the wall. In contrast, our model introduces a force directing individuals toward the a desired location, the center of the flock. This means a schooling flock, while temporarily moving away, would eventually return toward the selected point. Moreover, the anchor acts as an unbiased attraction force, causing individuals to turn either clockwise or counterclockwise. During this turning, the flock depolarizes, leading to phase changes. As a result, any observed schooling is temporary.

The delay's effect Another significant factor contributing to the expansion of milling is the delay. A delay between recognition and actuation has been documented to influence collective motion. **ZHOU2024delay**; Mijalkov et al. 2016; Holubec et al. 2021; Cavagna et al. 2017. If the anchor interacts with the flock with an attractive force, the delays does it with an alignment force. Because the delay, the particles tend to align with a past velocity. In other words, the polarization perceived by the particles is delayed, which leads to longer convergence on the 2 coordinated phases, only schooling or milling. This longer convergence time that the coordinated phase takes to stabilize creates opportunities for other phases to replace the current coordinated phase. Particularly the swarming phase expands because the flock itself maintains a uncoordinated cohesion on high attraction or low alignment (Fig. 3.8d). For the case of the bistability region, it expands on the borders of the coordinated phases, as the delay causes the flock to oscillate between milling and schooling phases more often.

Comparison between phase diagrams The comparison between the robotic (Fig. 3.9 a)) and the simulated phase diagrams (Fig. 3.8 d)) only differs for the small area of polarized bistability. The low-attraction region differs because collisions are more relevant in the experimental case than in simulation, where avoidance was included only in the attraction mechanism (low attraction equals low avoidance). In the high-attraction region, the small difference is attributable to the significant noise present in the alignment rule, caused by the compounding effect of optic flow calculations based on retinal location and apparent distance. All three primary effects of the robotic implementation expands the area of a stable milling phase.

3.6. Conclusion

Regardless of these implementation challenges, the collective motion model applied on the robotic platform reproduces several collective patterns, such as swarming, milling, and a bistable phase. To the best of our knowledge, for the first time, a visual flocking implementation using spherical robots produces a phase diagram

that matches almost perfectly with that of the simulation counterpart using the same visual model. Such an implementation bridges the gap between collective model simulation and real-world experiments, allowing for a framework independent of the N individuals belonging to the flock.

Future Work For future work, it would be interesting to implement an obstacle avoidance rule based on optic flow. With the divergence of the optic flow, potential collisions could be better avoided. It would also be valuable to explore how the presence of the most influential retinal object could affect the phase distribution or evaluate the current model to identify the average location of the most influential retinal object. Another interesting approach would be to introduce a true blind spot where there is no sensing. Potential applications include the evaluation of a moving anchor or transforming the concept of the anchor into a “guarding dog" individual—one individual that moves around and keeps the flock in the desired position—or to include the anchor as a static individual. Another interesting path would be the implementation to 20 or 40 individuals, or even incorporating a mixed-reality option with virtual individuals that interact with the robots to create flocks of hundreds.

Acknowledgements

We would also like to thank the two reviewers for their valuable suggestions and comments.
The participation of D.C. in this research was made possible by joint PhD grant from Aix-Marseille University.
D.C. and F.R. were also supported by Aix Marseille University and the CNRS (Life Science, Information Science, and Engineering and Science & technology Institutes).
The facilities for the experimental tests has been mainly provided by ROBOTEX 2.0 (Grants ROBOTEX ANR-10-EQPX-44-01 and TIRREX ANR-21-ESRE-0015).

Data Availability

Data acquired during the experiments and simulations are available online Castro et al. 2024a

Video Accessibility

Links to supplemental videos are provided in the supplementary information and can also be found at: youtube or Castro et al. 2024a.

Author contributions statement

D.C. performed the simulations, the robotic experiments, and data consolidation. C.E. & F.R. outlined the robot in the loop experiment and supervised its implementation. All authors: Proofread the manuscript.

3.7. Supplementary Information

3.7.1. Spheros' dynamical response and time delay τ estimation

The robotic setup is composed of several steps, from setting a heading and velocity, to recognizing a XY-coordinate and heading from the image. Each step involves their own input treatment, response time and processing. As such, we performed a step response analysis for both heading and velocity independently. This analysis allows to see if there are significant changes from the innate response of the simulation.

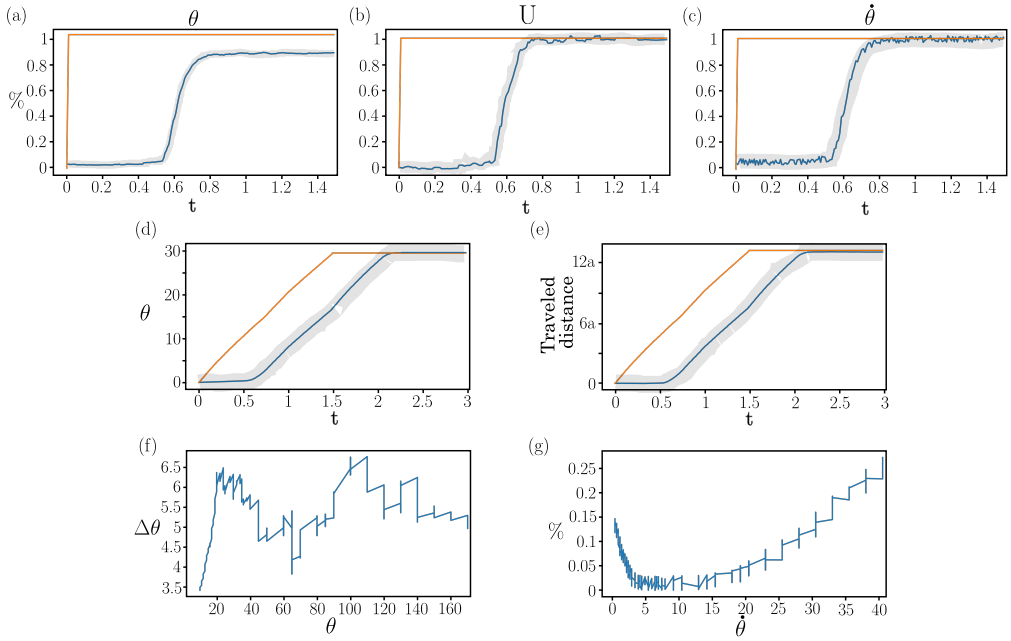

Figure 3.10.: Spheros' dynamical response
Spheros' dynamical response (The grey shade is the average of normalized standard deviation) a) Average normalized heading θ step response of the 520 experiments. b) Average normalized velocity U step response [a/sec] of the 220 experiments. c) Average normalized yaw $\dot{\theta}$ step response [degrees] of the 220 experiments. d) Average heading θ ramp response [degrees/sec] of the 220 experiments. e) Average velocity step response as seen in traveled distance (with $a = 3.5cm$) of the 220 experiments. f) Heading θ measurement average sensitivity in degrees of the 520 experiments. g) Yaw $\dot{\theta}$ measurement average sensitivity in percentage of the 520 experiments.

Fig. 3.10 a), shows the average step response for a change in the heading of the robot. The change ranged from 1 to 180 degrees. Fig. 3.10 b), shows the average step response for a change in the velocity of the robot. The change ranged from 1 to 25 a/sec. Fig. 3.10 c), shows the average step response for a change in the yaw $\dot{\theta}$ of the robot. Fig. 3.10 d)-e), shows this speeds ramps in the measured variables. All steps responses were performed 5 times with 10 different robots. The Y-axis is a percentage of the desired

heading. Finally, Fig. 3.10 f), shows the measurement sensitivity to changes in heading and Fig. 3.10 g), shows the error in following a desired yaw $\dot{\theta}$. These results show that the system has a median time delay τ of approximately 600ms in responding to any input signal change. The image processing takes from 10-50ms, the image acquisition take from 25-50ms. Leaving at best 400ms delay to the communication between the control and the robot, the inner processing and the actuation of the input. Therefore, the visual input for the collective motion model must reflect this transport delay.

3.7.2. Direct comparison between Supplemented visual model and robot-in-the-loop Experiments

The difference between Phase Diagrams constructed from simulated and robots-in-the-loop experiments for 10 individuals is very marginal Fig. 3.11. The mean squared error – MSE – between the phase diagram images of robots-in-the-loop and supplemented visual model is 2.48%.

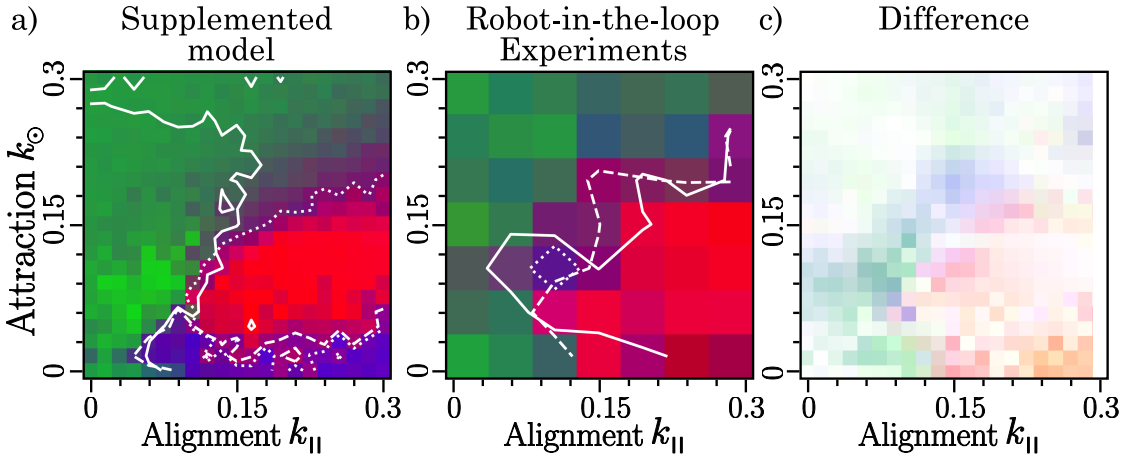

Figure 3.11.: Difference between the Phase Diagrams.
Difference between the Phase Diagrams. a) Supplemented visual model (10 individuals), b) Phase Diagram of Spheros' robots-in-the-loop (10 robots), c) Raw difference image difference between both phase diagram. The mean squared error – MSE – between the phase diagram images of robots-in-the-loop and supplemented visual model is 2.48%.

4. Conclusion

Contents

4.1. Summary 111
4.2. Contribution 117
4.3. Future Work 118

4.1. Summary

Collective motion is a widely observed phenomenon, characterized by the coordinated movement of individuals. From bacterial colonies and insect swarms to bird flocks and mammalian herds, this coordination emerges from simple interactions without any centralized control. The ubiquity of collective motion across vastly different species and environmental conditions suggests the existence of universal principles guiding these behaviors.

These behaviors have been extensively studied. Early models, such as Reynolds' Boids and the Vicsek model, introduced frameworks for collective motion by assuming individuals behave as self-propelled particles (SPP) that adhere to one or more basic behavioral rules: attraction, alignment, and avoidance. Each individual, or particle, considers the positions and velocities of its neighbors to influence its own movement. Through these individual actions, the models successfully replicated observed behaviors like swarming, schooling, and milling. However, many of these models assume omniscient awareness of neighbors' positions and velocities, a simplification that diverges significantly from the sensory-limited capabilities of real organisms. This discrepancy between traditional collective motion models and biologically plausible sensory constraints prompted the primary questions of this thesis:

How can collective motion be explained from a biologically plausible perspective? Does a visual collective motion model need to disregard traditional perspectives?

Visual model formulation The presented questions highlight a critical gap in current collective motion models: their reliance on abstract, distance-based metrics that assume idealized sensing capabilities. Chapter 2 addresses this limitation by developing a model that incorporates natural visual perception to replicate collective behaviors, while maintaining the traditional formulation to facilitate comparison.

The first step was to establish a representation of the visual field, as the selection of the representation would define all available visual cues. The representation selected was a binary panorama, a binary function that maps the presence of objects within

the projection of the visual field. A binary panorama projection encompasses all visible individuals, accounting for occlusion of distant individuals and the aggregation of partially occluded individuals into a single retinal object. The early visual cues derived from this representation include retinal position, optical size, and apparent distance (optical size function). Additionally, optic flow was computed by analyzing the evolution of retinal position and optical size over a single time step for each retinal object.

Implication of the selected visual representation The binary panorama and its associated visual cues enable an individual to react to the entire visible flock, as opposed to being restricted to a localized area around it. This expanded area of effect contrasts with traditional collective motion models, which enforce behavioral rules based on interactions with a subset of the flock. The increased information availability was an unforeseen consequence of implementing the visual model. To address this, Chapter 2 also introduces an omniscient model, which adheres to the same behavioral rules as traditional models but considers all individuals within the flock.

With a visual field representation in place, the next step was to identify the most suitable visual cues for implementing collective motion rules analogous to the Boids' model (attraction, alignment, avoidance). While optic flow and retinal position were clear choices for alignment and spatial awareness, determining the best cue to incorporate distance was more nuanced. Two candidates were considered: apparent distance and optical size.

Comparison of the available visual cues Optical size has a non-linear, negative exponential relationship with distance, with its maximum value occurring at touching distance (where the optical size is $\pi/2$). This relationship results in a disproportionate influence of nearby objects, particularly those within three body lengths. In contrast, apparent distance provides a linear effect, maintaining proportional influence across greater distances. The higher influence of larger retinal objects caused a subdivision of the flock, as illustrated in Fig. 2.8. This grouping effect is attributed to the inverse relationship with distance, a phenomenon also observed in the omniscient model under high alignment gains (Fig. A.7). In both cases, the dominance of closer objects skews the individual's perception of the flock's center of mass, creating a localized minimum.

While this localized minimum mimics the area-of-effect behavior of traditional models, it differs in stability. In traditional models, the locality constantly shifts, ultimately converging near the flock's true center of mass. In contrast, the visual and omniscient models maintain a stable local minimum relative to each individual. This distinction is evident in Fig. A.9, where the omniscient model shows a consistent flock spread across different flock sizes, unlike the visual model (Fig .A.2), which maintains constant density.

Conclusion on the visual model for collective motion Chapter 2 introduced a visual model for collective motion, utilizing **retinal position**, **apparent distance**, and **optic flow** as bio-plausible visual cues. Attraction was implemented using retinal position and apparent distance, while alignment relied on a combination of all three cues. Additionally, an omniscient model was presented, using the exact positions and velocities of all individuals in the flock. Both models successfully reproduced collective behaviors such as swarming, schooling, and milling. Moreover, they adhered to the same behavioral rules as traditional self-propelled particle (SPP) models, enabling seamless comparisons.

With the visual model successfully explaining collective motion from a biologically plausible perspective and aligning effortlessly with traditional models, the following complementary questions can be answered:

How can self-propelled particle simulations be efficiently modified to achieve a better alignment between simulation and robotic experimentation?

How can robotic experimentation can be simplified to introduce more individuals and still align with simulation?

Selecting the robots Finding answers to these questions has led to the development of multiple research robotic platforms, as shown in Fig. 1.24. All robotic platforms were developed with specific research questions in mind, and as such, each is equipped with the best available resources to address those questions. Because research questions vary, each robotic platform is unique, and no widely commercially available solution exists.[1] Specifically, there is no commercially available platform that fits a particle-like environment—in other words, one with holonomic movement. In the absence of such a solution, there are two primary options: either develop a custom platform (as most researchers opt to do) or adapt an existing solution to meet the specific requirements of the task. There were 3 main constrains that guided the selection process:

1. A self-propelled particle: The robot must move as closely it could to the particles presented on Chapter 2.

2. Solution availability: The robot should require the least amount of changes to answer the research question.

3. Flock size implementation: The robotic solution must be able to accommodate at least 10 individuals with the option to increase this number without a mayor redesign of the framework.

The first and most important constraint led to the discovery of an educational platform, Sphero. Sphero is a robotics company that developed and commercialized sphere-like robots for use in STEAM (Science, Technology, Engineering, Arts, and Mathematics) programs. The range of robots available extends from single-moving

[1] Kilobots and e-puck were considered, but their movement constraints ruled them out. HeRo 2.0 met the criteria but was excluded due to flock size limitations.

robots to LEGO®-style robotics. One of their solutions is called Sphero BOLT. The BOLT is a sphere-like robot with a transparent shell that moves by rotating the outer shell using an internal set of motors that maintain constant contact with the shell (similarly to a hamster in a hamster wheel). Inside the shell, it houses a complex, sealed, and reliable system controlled using Bluetooth technology and the Sphero educational app.

The Sphero BOLT is a robot that moves as desired but is a closed platform, meaning that integrating vision onto the robot was impossible. In addition, controlling several robots simultaneously or integrating the visual model was unexplored. Because the robots are designed for educational purposes, the communication protocol was reverse-engineered, and an open library is available for ROS, Python, and C++. These libraries cleared the path to integrate the control of the robots into the most suitable location taking into account the variable flock size and integration with the visual model.

Including vision with the Sphero BOLTs The last issue was to include vision in the system. This issue could not be solved by an internal or robot-by-robot option, as each robot is sealed. Thus, the challenge was how to use an external tool to integrate vision into the system. At this point, our choice of visual field representation once again simplified the problem. By using an external recognition system, each robot's panoramic binary map can be reproduced. This solution integrates the visual representation that the visual model presented in Chapter 2 employs.

The final system consisted of an external recognition tool (an overhead camera), the visual model for each robot, and the associated robot. The robot moves and behaves like a particle. The required development focused on robot recognition, model integration, and robot control. All three steps are independent of the number of robots involved in an experiment, allowing the solution to not be bound in terms of flock size capabilites.

The first issue, robot recognition, was addressed using the lights on top of each BOLT, which shine brightly through the transparent shell. By leveraging the three light sources, a unique pattern was designed. This pattern enabled clear segmentation of the robot and facilitated zooming in to directly measure its heading. However, recognizing multiple robots simultaneously posed a greater challenge. The same advantage of a transparent case, which enhances the visibility of the lights, creates a significant issue when multiple robots are in contact. In such cases, distinguishing the lights of robot A from those of robot B becomes difficult.

The effect of noise on the implementation Another challenge was the increased noise in calculating the position and heading of each robot. These values are critical because they are used to derive the retinal position, apparent distance, and optic flow of each robot. Any added noise compounds throughout the model. While the simulation of the visual model accounted for noise—attributable in the experimental context to heading measurement noise—experimental noise introduced additional

interference. This interference affected not only the apparent distance but also, to a greater extent, the optic flow.

The robotic simulation included a simple repulsion mechanism to avoid collisions, the constraints of delay on actuation, and an anchor. Unlike other robotic implementations of collective behaviors, we did not include a model of a world boundary. Instead, we opted to include an anchor. This decision was based on the foundational principles of the model. Modeling a wall from a visual point of view would involve representing a featureless background as a repulsive force. While this makes sense from a physics perspective, it is less intuitive from a visual one. Instead, the focus was on incorporating a feature that tied the particles to a specific location: an anchor. Furthermore, this anchor is not restricted to being virtual, as proposed in Chapter 3. It could just as easily be represented by a marker or a terrain feature.

The effect of the anchor The anchor also exhibited an unforeseen behavior: the favoritism of milling over schooling. A world boundary, in most cases, interacts with particles in a way that encourages "bumpy" schooling along the wall. Schooling is a translational phase, observed during the migration or movement of a flock. If the model incorporates a wall, schooling would inevitably interact with it, causing the school-wall interaction to become bouncy and give rise to this characteristic behavior.

In the presented model, there was no wall, nor any direct interaction with the "world edges." Instead, the model generated an attractive force towards the center of the area of interest. This meant that a schooling flock would, at some point, move away from the center but eventually return towards it. Due to the unbiased nature of the attraction rule, each individual could turn either clockwise or counterclockwise to return to the center. This turning probability caused the flock to depolarize during turning, leading to the emergence of a new phase. As a result, any schooling observed was temporary. This discussion explains the lack of schooling in the model with an anchor. However, why does adding the anchor favor milling? Here, the special stable state of milling becomes relevant.

Fig. 2.10 shows the statistical distribution of a stable milling phase, aligning with Fig. 1.13 (Papadopoulou et al. 2023b). Both figures demonstrate that milling is a phase primarily dependent on attraction. Low proportions of alignment assist with the directionality of the flock during milling.

Milling as the main stable phase The anchor acts as an attractive force toward a single location. It has its own strength but, at the right distance, its effect is equivalent to that of any other individual within the flock. Additionally, the cyclical nature of the flock under the influence of the anchor creates several opportunities for achieving a stable distance. This distance is radial, meaning that the path taken by a particle to maintain a constant distance is circular and centered around the anchor. If more particles engage in the same path, milling emerges. The inclusion of the anchor creates conditions conducive to this behavior, which is why milling is favored when the anchor is added.

Although the anchor greatly influences the milling phase, another factor also contributes: delay. A delay between recognition and action has been documented to influence collective motion. If the anchor stabilizes milling by expanding the stable attraction space, the delay has a similar effect on the alignment space. The non-immediacy introduced by the delay causes particles to align with a previous velocity. In other words, achieving full polarization of the flock takes longer. The extra time required for full polarization creates opportunities for the attractive stable state of milling to stabilize.

Furthermore, in a stable milling phase, the angular velocity of each individual remains constant or nearly constant, meaning that the alignment component never reaches zero—a condition amplified by the delay. Together, the anchor and the delay expand the stable region of milling while reducing the stable region of schooling. Fig. 4.1 illustrates the expansion of the simulated phase diagram with twice the original alignment.

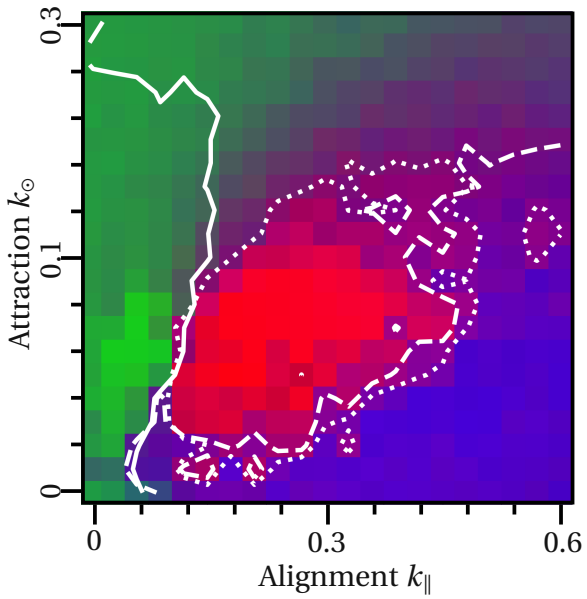

Figure 4.1.: Expanded phase diagram of the visual model that included the anchor, the delay and the avoidance. The original phase diagram only explored the alignment term ($k_\parallel$) from *0* to *0.3*, the expanded phase diagram explores the alignment term from *0* to *0.6*, double the original values.

The expanded phase diagram shows a large bistable zone, but this bistability tends to be highly polarized (the shade of purple leans more toward blue than red). Although it does not represent a stable schooling phase, these results support the discussion above. The inclusion of both anchor and delay reduces the flock's polarization.

The comparison between the experimental phase diagram and the simulated phase diagram mainly differs in areas of polarized bistability. The low-attraction region differs because collisions are more relevant in the experimental case than in simulation, where avoidance was included only in the attraction mechanism (low attraction

equals low avoidance). In the high-attraction region, the difference is attributable to the significant noise present in the alignment rule, caused by the compounding effect of optic flow calculations based on retinal location and apparent distance. All three primary effects of the robotic implementation negatively influence flock polarization.

Implementation conclusion Despite these challenges, the implementation resulted in an environment that aligned well with minimal-effort simulations and real-world patterns. Moreover, the reciprocity of the results is not limited to specific cases but extends to the phase diagram's distribution. No other robotic collective motion implementation reproduces all the phases that the numerical simulation replicated.

The proposed solution establishes an environment for utilizing SPP (Self-Propelled Particle) collective motion models with real-world robotics. This environment was also designed from the outset to be bio-plausible. In other words, this study provides tools to describe natural phenomena using biological insights, physics-based techniques, and an engineering platform, harmoniously integrating all three approaches.

4.2. Contribution

This work began with four research questions that can be summarized into two main topics: bio-plausible modeling and the robotic implementation of collective motion.

The modeling topic addresses the first two research questions. The first question focuses on developing a model in isolation. However, addressing this question alone leads to a solution that has already been explored—constructing a completely new system to describe collective motion while disregarding the extensive history of the subject. This limitation is addressed by the second question, which aims to integrate this rich historical foundation into the model.

The first major contribution of this work is a visual model of collective behavior—a model that replicates the traditional phases of collective motion using cues observed in nature. These cues serve as the foundation of the well-established, 40-year-old rules prevalent in the literature, such as those proposed by Vicsek, Reynolds, and Couzin. By explaining these rules through natural visual cues, the work naturally connects traditional collective motion models with biological descriptors, specifically visual stimuli.

The implementation topic encompasses the latter two research questions. Both questions target the same overarching goal: to improve the implementability of the model within an artificial flock of robots. However, they approach this goal via two contrasting pathways—one through complex simulations that enhance the alignment of virtual models with real-world dynamics, and the other by modifying the robots to more closely resemble their virtual counterparts.

Following the example of most studies, this work pursues both pathways simultaneously. Modifications were incorporated into the model to better reflect experimental conditions —such as avoidance, delay, and anchor— while the selection of robots was tailored to match the assumptions of particle models.

The result is a simulation and experimental setup that are closely aligned. Furthermore, the model was made dimensionless by selecting a length scale equal to the robot's radius and a time scale based on the time required for a robot to traverse its radius. This approach ensures that all simulations correlate to any speed setting of the robots.

Consequently, the supplemented model and the robotic platform can be used independently, but their integration creates a robust environment for studying collective motion. This environment incorporates bio-plausible information, applies traditional concepts from particle physics, and operates on a flexible robotic platform.

In summary, the contributions of this thesis are as follows:

- The development of a visual model for collective motion based exclusively on visual cues, specifically retinal position, apparent distance, and optic flow.
- The implementation of this visual model within the framework of traditional self-propelled particle (SPP) models, such as those proposed by Vicsek, Reynolds, and Couzin.
- The design and development of a robotic platform that closely mimics the behavior modeled by self-propelled particle assumptions.
- The successful implementation of the visual model in a robots-in-the-loop approach, which effectively reproduced multiple collective behaviors.
- The development of a framework for collective motion research that balances traditional physics models, biological information and robotic experimentation. With high experimental fidelity.

4.3. Future Work

For future work, there are two main routes to take. The modeling path and the robotic path. From the modeling path, a more complex visual model can be explored. The complexity can be from the early vision point of view to handle a grayscale visual field (so more features could be extracted), a true blind spot with discontinuities or something like an image-based visual field (more in line with cameras). Another very interesting point would be to change the model to include particles of different sizes or speeds, to have the noise to be added on the measurement of the vision instead of being on the actuation or to have a model that includes changes on the linear velocities of the particles. Finally, to implement the model on 3D, which could be with a spherical coordinate system and have control on angular velocities on two angles. Or to have a cylindrical model and decouple angular velocities and altitude.

From the robotic path, the most important step should be to include a more complex avoidance rule that ensures no collision, from both the early vision point of view and optic flow. Then there are two main options, increase the flock size (to 30, 50 or 100) or improve each individual. The first point would be to have the same robots but to add

more as they are easily scalable. Then for this purpose the recognition of each of the robots by the overhead camera needs to be improved. Following the communication protocol between the control loop and the robot, that should be done on a Linux-based operating system (a windows based operating system does not allow for the scalability past what the current work presented). To improve each robot, then a custom hardware should be designed or adapted. One that allows to have an onboard camera, which would avoid needing a centralized hub for recognition and a holonomic movement. One solution would be a Raspberry Pi or ESP32 board with triangular omnidirectional wheel configuration. Using this kind of open development boards allows having a camera and an optic flow sensor but also require a long development time and effort.

Bibliography

[Alb96] Ezequiel V Albano. "Self-organized collective displacements of self-driven individuals". In: *Physical review letters* 77.10 (1996), p. 2129 (cit. on p. 69).

[AB02] Réka Albert et al. "Statistical mechanics of complex networks". In: *Reviews of modern physics* 74.1 (2002), p. 47 (cit. on p. 50).

[Aok84] Ichiro Aoki. "Internal Dynamics of Fish Schools in Relation to Inter-fish Distance". In: *Bulletin of the Japanese Society of Scientific Fisheries* 50.5 (1984), pp. 751–758 (cit. on p. 70).

[Apl+14] Lucy M Aplin et al. "Individual-level personality influences social foraging and collective behaviour in wild birds". In: *Proceedings of the Royal Society B: Biological Sciences* 281.1789 (2014), p. 20141016 (cit. on p. 23).

[Ari+14] Gil Ariel et al. "Individual pause-and-go motion is instrumental to the formation and maintenance of swarms of marching locust nymphs". In: *PloS one* 9.7 (2014), e101636 (cit. on p. 70).

[Att+14] Alessandro Attanasi et al. "Information transfer and behavioural inertia in starling flocks". In: *Nature physics* 10.9 (2014), pp. 691–696 (cit. on p. 23).

[BA08] Gabriel Baglietto et al. "Finite-size scaling analysis and dynamic study of the critical behavior of a model for the collective displacement of self-driven individuals". In: *Physical Review E—Statistical, Nonlinear, and Soft Matter Physics* 78.2 (2008), p. 021125 (cit. on p. 42).

[Bak+81] Myron Charles Baker et al. "Foraging success in junco flocks and the effects of social hierarchy". In: *Animal Behaviour* 29.1 (1981), pp. 137–142 (cit. on p. 23).

[Bal+08a] Michele Ballerini et al. "Empirical investigation of starling flocks: a benchmark study in collective animal behaviour". In: *Animal behaviour* 76.1 (2008), pp. 201–215 (cit. on pp. 21, 70).

[Bal+08b] Michele Ballerini et al. "Interaction ruling animal collective behavior depends on topological rather than metric distance: Evidence from a field study". In: *Proceedings of the national academy of sciences* 105.4 (2008), pp. 1232–1237 (cit. on pp. 25, 37, 38, 69).

[BP16] Lucas Barberis et al. "Large-scale patterns in a minimal cognitive flocking model: incidental leaders, nematic patterns, and aggregates". In: *Physical review letters* 117.24 (2016), p. 248001 (cit. on p. 69).

[BR20] Renaud Bastien et al. “A model of collective behavior based purely on vision”. In: *Science advances* 6.6 (2020), eaay0792 (cit. on pp. 65, 66, 70, 90).

[Bat15] Stephen Battersby. “The cells that flock together”. In: *Proceedings of the National Academy of Sciences* 112.26 (2015), pp. 7883–7885 (cit. on p. 22).

[Baz+08] Sepideh Bazazi et al. “Collective motion and cannibalism in locust migratory bands”. In: *Current biology* 18.10 (2008), pp. 735–739 (cit. on p. 31).

[Baz+10] Sepideh Bazazi et al. “The social context of cannibalism in migratory bands of the mormon cricket”. In: *PLoS One* 5.12 (2010), e15118 (cit. on p. 31).

[Baz+11] Sepideh Bazazi et al. “Nutritional state and collective motion: from individuals to mass migration”. In: *Proceedings of the Royal Society B: Biological Sciences* 278.1704 (2011), pp. 356–363 (cit. on p. 21).

[Bea03] Guy Beauchamp. “Group-size effects on vigilance: a search for mechanisms”. In: *Behavioural processes* 63.3 (2003), pp. 111–121 (cit. on p. 37).

[Bec+06] Ch Becco et al. “Experimental evidences of a structural and dynamical transition in fish school”. In: *Physica A: Statistical Mechanics and its Applications* 367 (2006), pp. 487–493 (cit. on p. 25).

[Ber+21] Lucia Bergantin et al. “Oscillations make a self-scaled model for honeybees’ visual odometer reliable regardless of flight trajectory”. In: *Journal of the Royal Society Interface* 18.182 (2021), p. 20210567 (cit. on pp. 70, 90).

[Bia+12] William Bialek et al. “Statistical mechanics for natural flocks of birds”. In: *Proceedings of the National Academy of Sciences* 109.13 (2012), pp. 4786–4791 (cit. on p. 90).

[Bod+10] Nikolai WF Bode et al. “How perceived threat increases synchronization in collectively moving animal groups”. In: *Proceedings of the Royal Society B: Biological Sciences* 277.1697 (2010), pp. 3065–3070 (cit. on p. 69).

[BFW11] Nikolai WF Bode et al. “Limited interactions in flocks: relating model simulations to empirical data”. In: *Journal of The Royal Society Interface* 8.55 (2011), pp. 301–304 (cit. on p. 90).

[BWF11] Nikolai WF Bode et al. “Social networks and models for collective motion in animals”. In: *Behavioral ecology and sociobiology* 65.2 (2011), pp. 117–130 (cit. on p. 40).

[Bon+97] Eric Bonabeau et al. “Self-organization in social insects”. In: *Trends in ecology & evolution* 12.5 (1997), pp. 188–193 (cit. on p. 23).

[BDT99] Eric Bonabeau et al. *Swarm Intelligence: From Natural to Artificial Systems*. Oxford University Press, Oct. 1999. ISBN: 9780195131581. DOI: 10.1093/oso/9780195131581.001.0001. URL: https://doi.org/10.1093/oso/9780195131581.001.0001 (cit. on p. 20).

[Bon23] Kathryn Bonnen. “Motion vision: Fish swimming to see”. In: *Current Biology* 33.1 (2023), R30–R32 (cit. on pp. 21, 70, 90).

[Bou16] Roland Bouffanais. *Design and control of swarm dynamics*. Vol. 1. Springer, 2016 (cit. on p. 21).

[BB91] Elena O Budrene et al. “Complex patterns formed by motile cells of Escherichia coli”. In: *Nature* 349.6310 (1991), pp. 630–633 (cit. on p. 23).

[Buh+06] Jerome Buhl et al. “From disorder to order in marching locusts”. In: *Science* 312.5778 (2006), pp. 1402–1406 (cit. on pp. 23, 29).

[Cal+14] Daniel S Calovi et al. “Swarming, schooling, milling: phase diagram of a data-driven fish school model”. In: *New journal of Physics* 16.1 (2014), p. 015026 (cit. on pp. 20, 47, 72, 75, 76, 100).

[Cam95] TA Camus. “Calculating time-to-contact using real-time quantized optical flow”. In: (1995) (cit. on p. 63).

[CS05] Gilles Caprari et al. “Mobile micro-robots ready to use: Alice”. In: *2005 IEEE/RSJ international conference on intelligent robots and systems*. IEEE. 2005, pp. 3295–3300 (cit. on p. 51).

[Car81] Thomas Caraco. “Risk-sensitivity and foraging groups”. In: *Ecology* 62.3 (1981), pp. 527–531 (cit. on p. 23).

[CMP80] Thomas Caraco et al. “Avian flocking in the presence of a predator”. In: *Nature* 285.5764 (1980), pp. 400–401 (cit. on p. 23).

[CER24a] Diego Castro et al. *Data and Code of: Visual collective behaviors on spherical robots*. Sept. 2024. DOI: 10.17605/OSF.IO/UGWR9. URL: osf.io/ugwr9 (cit. on p. 107).

[CRE24] Diego Castro et al. “Modeling collective behaviors from optic flow and retinal cues”. In: *Phys. Rev. Res.* 6 (2 Apr. 2024), p. 023016. DOI: 10.1103/PhysRevResearch.6.023016. URL: https://link.aps.org/doi/10.1103/PhysRevResearch.6.023016 (cit. on pp. 68, 90–92, 95–97, 102–104, 164).

[CER24b] Diego Castro et al. *Source Code for the Flocking by Eye Project*. Apr. 2024. URL: https://gitlab.com/FlokingByEyeDiego/ (cit. on p. 83).

[Cau+14] Jean-Baptiste Caussin et al. “Emergent spatial structures in flocking models: a dynamical system insight”. In: *Physical review letters* 112.14 (2014), p. 148102 (cit. on pp. 69, 72).

[Cav+08a] Andrea Cavagna et al. “The STARFLAG handbook on collective animal behaviour: Part I, empirical methods”. In: *arXiv preprint arXiv:0802.1668* (2008) (cit. on pp. 25, 38).

[Cav+08b] Andrea Cavagna et al. “The STARFLAG handbook on collective animal behaviour: Part II, three-dimensional analysis”. In: *arXiv preprint arXiv:0802.1674* (2008) (cit. on p. 38).

[Cav+10] Andrea Cavagna et al. “Scale-free correlations in starling flocks”. In: *Proceedings of the National Academy of Sciences* 107.26 (2010), pp. 11865–11870 (cit. on p. 38).

[Cav+13] Andrea Cavagna et al. “Diffusion of individual birds in starling flocks”. In: *Proceedings of the Royal Society B: Biological Sciences* 280.1756 (2013), p. 20122484 (cit. on p. 23).

[Cav+15] Andrea Cavagna et al. “Flocking and turning: a new model for self-organized collective motion”. In: *Journal of Statistical Physics* 158 (2015), pp. 601–627 (cit. on p. 90).

[Cav+17] Andrea Cavagna et al. “Dynamic scaling in natural swarms”. In: *Nature Physics* 13.9 (2017), pp. 914–918 (cit. on p. 106).

[Cha+09] Arthur D Chapman et al. “Numbers of living species in Australia and the world”. In: (2009) (cit. on p. 27).

[CT19] Henry J Charlesworth et al. “Intrinsically motivated collective motion”. In: *Proceedings of the National Academy of Sciences* 116.31 (2019), pp. 15362–15367 (cit. on p. 75).

[Cha+08] Hugues Chaté et al. “Collective motion of self-propelled particles interacting without cohesion”. In: *Phys. Rev. E* 77 (4 Apr. 2008), p. 046113. DOI: 10.1103/PhysRevE.77.046113. URL: https://link.aps.org/doi/10.1103/PhysRevE.77.046113 (cit. on p. 90).

[CSH16] Bertrand Collignon et al. “A stochastic vision-based model inspired by zebrafish collective behaviour in heterogeneous environments”. In: *Royal Society open science* 3.1 (2016), p. 150473 (cit. on pp. 65, 70, 90).

[CH18] A Costanzo et al. “Spontaneous emergence of milling (vortex state) in a Vicsek-like model”. In: *Journal of Physics D: Applied Physics* 51.13 (2018), p. 134004 (cit. on p. 86).

[Cou+02] Iain D Couzin et al. “Collective memory and spatial sorting in animal groups”. In: *Journal of theoretical biology* 218.1 (2002), pp. 1–11 (cit. on pp. 21, 43, 45, 90).

[Cro+14] Thomas W Cronin et al. *Visual ecology*. Princeton University Press, 2014 (cit. on p. 55).

[CSB65] John M Cullen et al. “Methods for measuring the three-dimensional structure of fish schools”. In: *Animal Behaviour* 13.4 (1965), pp. 534–543 (cit. on p. 25).

[CMV01] Andras Czirók et al. “Theory of periodic swarming of bacteria: Application to Proteus mirabilis”. In: *Physical Review E* 63.3 (2001), p. 031915 (cit. on p. 26).

[Czi+96] András Czirók et al. “Formation of complex bacterial colonies via self-generated vortices”. In: *Physical Review E* 54.2 (1996), p. 1791 (cit. on p. 26).

[Das+24] Suchismita Das et al. “Flocking by turning away”. In: *Physical Review X* 14.3 (2024), p. 031008 (cit. on p. 24).

[Dav+21] Jacob D Davidson et al. “Collective detection based on visual information in animal groups”. In: *Journal of the Royal Society Interface* 18.180 (2021), p. 20210142 (cit. on pp. 69, 90).

[De +22] Guido CHE De Croon et al. “Accommodating unobservability to control flight attitude with optic flow”. In: *Nature* 610.7932 (2022), pp. 485–490 (cit. on pp. 70, 90).

[DBD16] Johann Delcourt et al. “Collective vortex behaviors: Diversity, proximate, and ultimate causes of circular animal group movements”. In: *The Quarterly review of biology* 91.1 (2016), pp. 1–24 (cit. on pp. 43–45, 86).

[Dev+06] Claire L Devereux et al. “Predator detection and avoidance by starlings under differing scenarios of predation risk”. In: *Behavioral Ecology* 17.2 (2006), pp. 303–309 (cit. on p. 56).

[Dor+13] Marco Dorigo et al. “Swarmanoid: a novel concept for the study of heterogeneous robotic swarms”. In: *IEEE Robotics & Automation Magazine* 20.4 (2013), pp. 60–71 (cit. on p. 51).

[DTT21] Marco Dorigo et al. “Swarm robotics: Past, present, and future [point of view]”. In: *Proceedings of the IEEE* 109.7 (2021), pp. 1152–1165 (cit. on p. 51).

[DD12] Ron Douglas et al. *The visual system of fish*. Springer Science & Business Media, 2012 (cit. on p. 57).

[Eic+11] Hubert Eichner et al. “Internal structure of the fly elementary motion detector”. In: *Neuron* 70.6 (2011), pp. 1155–1164 (cit. on p. 73).

[Eml52] John T Emlen. “Flocking behavior in birds”. In: *The Auk* 69.2 (1952), pp. 160–170 (cit. on p. 36).

[Etm+22] Amin Etminan et al. “Flow visualization: state-of-the-art development of micro-particle image velocimetry”. In: *Measurement Science and Technology* 33.9 (2022), p. 092002 (cit. on p. 24).

[ER15] Fabien Expert et al. “Flying over uneven moving terrain based on optic-flow cues without any need for reference frames or accelerometers”. In: *Bioinspiration & biomimetics* 10.2 (2015), p. 026003 (cit. on pp. 70, 90).

[Fer88] Russell D Fernald. “Aquatic adaptations in fish eyes”. In: *Sensory biology of aquatic animals*. Springer, 1988, pp. 435–466 (cit. on p. 57).

[Fer12] Esteban Fernández-Juricic. “Sensory basis of vigilance behavior in birds: synthesis and future prospects”. In: *Behavioural Processes* 89.2 (2012), pp. 143–152 (cit. on p. 57).

[FSK04] Esteban Fernández-Juricic et al. “Flock density, social foraging, and scanning: an experiment with starlings”. In: *Behavioral ecology* 15.3 (2004), pp. 371–379 (cit. on p. 23).

[Fil+18] Audrey Filella et al. “Model of collective fish behavior with hydrodynamic interactions”. In: *Physical review letters* 120.19 (2018), p. 198101 (cit. on pp. 72, 75, 100).

[Flo+13] Dario Floreano et al. “Miniature curved artificial compound eyes”. In: *Proceedings of the National Academy of Sciences* 110.23 (2013), pp. 9267–9272 (cit. on pp. 70, 90).

[FRL89] Nicolas Franceschini et al. “Directionally selective motion detection by insect neurons”. In: *Facets of vision*. Springer, 1989, pp. 360–390 (cit. on pp. 63, 73).

[FRS07] Nicolas Franceschini et al. “A bio-inspired flying robot sheds light on insect piloting abilities”. In: *Current Biology* 17.4 (2007), pp. 329–335 (cit. on pp. 70, 90).

[Fra+16] Nigel R Franks et al. “Social behaviour and collective motion in plant-animal worms”. In: *Proceedings of the Royal Society B: Biological Sciences* 283.1825 (2016), p. 20152946 (cit. on pp. 30, 31).

[GJT08] Jacques Gautrais et al. “Key behavioural factors in a self-organised fish school model”. In: *Annales Zoologici Fennici*. Vol. 45. 5. BioOne. 2008, pp. 415–428 (cit. on pp. 44, 45).

[Gau+09] Jacques Gautrais et al. “Analyzing fish movement as a persistent turning walker”. In: *Journal of mathematical biology* 58 (2009), pp. 429–445 (cit. on p. 45).

[Gau+12] Jacques Gautrais et al. “Deciphering interactions in moving animal groups”. In: (2012) (cit. on p. 46).

[Ger97] Mark Germine. “The physiology of collective consciousness”. In: *World Futures: Journal of General Evolution* 48.1-4 (1997), pp. 57–104 (cit. on p. 40).

[GH06] Mimi M Ghim et al. “Spatial contrast sensitivity of birds”. In: *Journal of Comparative Physiology A* 192 (2006), pp. 523–534 (cit. on p. 57).

[Gib50] James J Gibson. “The perception of the visual world.” In: (1950) (cit. on pp. 58, 59).

[GC10] Francesco Ginelli et al. “Relevance of metric-free interactions in flocking phenomena”. In: *Physical review letters* 105.16 (2010), p. 168103 (cit. on pp. 69, 70, 90).

[Gök08] Fatih Gökçe. “To flock or not to flock: pros and cons of flocking in long-range" migration" of mobile robot swarms”. MA thesis. Middle East Technical University, 2008 (cit. on p. 23).

[Gor96] Deborah M Gordon. “The organization of work in social insect colonies.” In: *Nature* 380.6570 (1996) (cit. on p. 27).

[Gor14] Deborah M Gordon. “The ecology of collective behavior”. In: *PLoS biology* 12.3 (2014), e1001805 (cit. on p. 40).

[GC04] Guillaume Grégoire et al. “Onset of collective and cohesive motion”. In: *Physical review letters* 92.2 (2004), p. 025702 (cit. on p. 42).

[GAJ08] D Grossman et al. “Emergence of agent swarm migration and vortex formation through inelastic collisions”. In: *New Journal of Physics* 10.2 (2008), p. 023036 (cit. on pp. 48, 49).

[HD02] Adam T Hayes et al. “Self-organized flocking with agent failure: Off-line optimization and demonstration with real robots”. In: *Proceedings 2002 IEEE international conference on robotics and automation (cat. no. 02ch37292)*. Vol. 4. IEEE. 2002, pp. 3900–3905 (cit. on p. 52).

[Hep74] Frank H Heppner. “Avian flight formations”. In: *Bird-banding* 45.2 (1974), pp. 160–169 (cit. on p. 36).

[Hol+05] Owen Holland et al. “Beyond swarm intelligence: the ultraswarm”. In: *Proceedings 2005 IEEE Swarm Intelligence Symposium, 2005. SIS 2005.* IEEE. 2005, pp. 217–224 (cit. on p. 53).

[Hol+21] Viktor Holubec et al. “Finite-size scaling at the edge of disorder in a time-delay Vicsek model”. In: *Physical review letters* 127.25 (2021), p. 258001 (cit. on p. 106).

[HS81] Berthold KP Horn et al. “Determining optical flow”. In: *Artificial intelligence* 17.1-3 (1981), pp. 185–203 (cit. on p. 62).

[Hor77] G Adrian Horridge. “The compound eye of insects”. In: *Scientific American* 237.1 (1977), pp. 108–121 (cit. on p. 55).

[Hug+18] Lacey F Hughey et al. “Challenges and solutions for studying collective animal behaviour in the wild”. In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 373.1746 (2018), p. 20170005 (cit. on p. 22).

[IU24] Susumu Ito et al. *Selective decision-making and collective behavior of fish by the motion of visual attention*. July 2024. DOI: 10.1093/pnasnexus/pgae264. eprint: https://academic.oup.com/pnasnexus/article-pdf/3/7/pgae264/58645783/pgae264.pdf. URL: https://doi.org/10.1093/pnasnexus/pgae264 (cit. on pp. 66, 90).

[JLM03] Ali Jadbabaie et al. “Coordination of groups of mobile autonomous agents using nearest neighbor rules”. In: *IEEE Transactions on automatic control* 48.6 (2003), pp. 988–1001 (cit. on p. 50).

[KSA17] Andreas Kaiser et al. “Flocking ferromagnetic colloids”. In: *Science advances* 3.2 (2017), e1601469 (cit. on p. 22).

[Kat+11] Yael Katz et al. “Inferring the structure and dynamics of interactions in schooling fish”. In: *Proceedings of the National Academy of Sciences* 108.46 (2011), pp. 18720–18725 (cit. on pp. 23, 69).

[KS71] Evelyn F Keller et al. “Model for chemotaxis”. In: *Journal of theoretical biology* 30.2 (1971), pp. 225–234 (cit. on p. 26).

[Ker13] Serge Kernbach. *Handbook of collective robotics: fundamentals and challenges*. CRC Press, 2013 (cit. on p. 52).

[KF69] K Kirschfeld et al. "A mechanism for controlling light flux in the rhabdomeres of the compound eye of Musca". In: *Cybernetics* 6 (1969), pp. 13–22 (cit. on p. 55).

[KD87] Jan J Koenderink et al. "Facts on optic flow". In: *Biological cybernetics* 56.4 (1987), pp. 247–254 (cit. on p. 60).

[KKL05] Sergey Kornienko et al. "Minimalistic approach towards communication and perception in microrobotic swarms". In: *2005 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE. 2005, pp. 2228–2234 (cit. on p. 51).

[KHH98] Holger G Krapp et al. "Dendritic structure and receptive-field organization of optic flow processing interneurons in the fly". In: *Journal of neurophysiology* 79.4 (1998), pp. 1902–1917 (cit. on p. 60).

[KL23] David L Krongauz et al. "Collective evolution learning model for vision-based collective motion with collision avoidance". In: *PLoS One* 18.5 (2023), e0270318 (cit. on pp. 66, 90).

[KAK24] David L Krongauz et al. "Vision-based collective motion: A locust-inspired reductionist model". In: *PLOS Computational Biology* 20.1 (2024), e1011796 (cit. on pp. 21, 65, 70, 90).

[Kud+08] Arshad Kudrolli et al. "Swarming and swirling in self-propelled polar granular rods". In: *Physical review letters* 100.5 (2008), p. 058001 (cit. on p. 24).

[KC09] VL Kulinskii et al. "The kinetic regime of the Vicsek model". In: *arXiv preprint arXiv:0910.5707* (2009) (cit. on p. 42).

[Laf+23] Baptiste Lafoux et al. "Illuminance-tuned collective motion in fish". In: *Communications Biology* 6.1 (2023), p. 585 (cit. on pp. 57, 58, 90).

[Laf+24] Baptiste Lafoux et al. "Confinement-driven state transition and bistability in schooling fish". In: *Physical Review E* 110.3 (2024), p. 034613 (cit. on pp. 35, 36, 103).

[Laz79] John Lazarus. "The early warning function of flocking in birds: an experimental study with captive quelea". In: *Animal Behaviour* 27 (1979), pp. 855–865 (cit. on p. 90).

[LAG09] Bertrand H Lemasson et al. "Collective motion in animal groups from a neurobiological perspective: the adaptive benefits of dynamic sensory loads and selective attention". In: *Journal of theoretical biology* 261.4 (2009), pp. 501–510 (cit. on p. 70).

[LAG13] Bertrand H Lemasson et al. "Motion-guided attention promotes adaptive communications during social navigation". In: *Proceedings of the Royal Society B: Biological Sciences* 280.1754 (2013), p. 20122003 (cit. on p. 69).

[LRC00] Herbert Levine et al. “Self-organization in systems of self-propelled particles”. In: *Physical Review E* 63.1 (2000), p. 017101 (cit. on p. 42).

[LL17] Benno Liebchen et al. “Collective behavior of chiral active matter: Pattern formation and enhanced flocking”. In: *Physical review letters* 119.5 (2017), p. 058002 (cit. on p. 69).

[Lin+24] Marcus Lin et al. “Emergent collective motion of self-propelled condensate droplets”. In: *Physical Review Letters* 132.5 (2024), p. 058203 (cit. on p. 23).

[Lop+12] Ugo Lopez et al. “From behavioural analyses to models of collective motion in fish schools”. In: *Interface focus* 2.6 (2012), pp. 693–707 (cit. on pp. 22, 34, 36).

[LZQ22] Xinbiao Lu et al. “An improved Vicsek model of swarm based on remote neighbors strategy”. In: *Physica A: Statistical Mechanics and its Applications* 587 (2022), p. 126553 (cit. on p. 75).

[LK81] Bruce D Lucas et al. “An iterative image registration technique with an application to stereo vision”. In: *IJCAI’81: 7th international joint conference on Artificial intelligence*. Vol. 2. 1981, pp. 674–679 (cit. on p. 62).

[LWG14] Enkeleida Lushi et al. “Fluid flows created by swimming bacteria drive self-organization in confined suspensions”. In: *Proceedings of the National Academy of Sciences* 111.27 (2014), pp. 9733–9738 (cit. on p. 27).

[MD78] Peter F Major et al. “The three-dimensional structure of airborne bird flocks”. In: *Behavioral Ecology and Sociobiology* 4 (1978), pp. 111–122 (cit. on p. 25).

[Mat94] Maja J Mataric. “Interaction and intelligent behavior”. In: (1994) (cit. on p. 52).

[Mat+19] David Mateo et al. “Optimal network topology for responsive collective behavior”. In: *Science advances* 5.4 (2019), eaau0999 (cit. on pp. 54, 69).

[Mez+24] David Mezey et al. “Purely vision-based collective movement of robots”. In: *arXiv preprint arXiv:2406.17106* (2024) (cit. on p. 90).

[Mij+16] Mite Mijalkov et al. “Engineering sensorial delay to control phototaxis and emergent collective behaviors”. In: *Physical Review X* 6.1 (2016), p. 011008 (cit. on p. 106).

[MG12] Noam Miller et al. “From schooling to shoaling: patterns of collective motion in zebrafish (Danio rerio)”. In: *PloS one* 7.11 (2012), e48865 (cit. on p. 35).

[MBM10] Shradha Mishra et al. “Fluctuations and pattern formation in self-propelled particles”. In: *Physical Review E* 81.6 (2010), p. 061916 (cit. on pp. 42, 69, 90).

[Mon+05] Francesco Mondada et al. “The cooperation of swarm-bots: Physical interactions in collective robotics”. In: *IEEE Robotics & Automation Magazine* 12.2 (2005), pp. 21–28 (cit. on p. 51).

[Mon+09] Francesco Mondada et al. “The e-puck, a robot designed for education in engineering”. In: *Proceedings of the 9th conference on autonomous robot systems and competitions*. Vol. 1. 1. IPCB: Instituto Politécnico de Castelo Branco. 2009, pp. 59–65 (cit. on p. 51).

[Mor+15] Alexandre Morin et al. “Collective motion with anticipation: Flocking, spinning, and swarming”. In: *Physical Review E* 91.1 (2015), p. 012134 (cit. on pp. 21, 72).

[MSC+08] Monica Moroni et al. “Mapping Starling Flocks through Image Analysis Techniques”. In: *14th Int. Symp. on Applications of Laser Tehniques to Fluid Mechanics*. 2008 (cit. on p. 22).

[NDV07] Máté Nagy et al. “New aspects of the continuous phase transition in the scalar noise model (SNM) of collective motion”. In: *Physica A: Statistical Mechanics and its Applications* 373 (2007), pp. 445–454 (cit. on p. 42).

[Nag+10] Máté Nagy et al. “Hierarchical group dynamics in pigeon flocks”. In: *Nature* 464.7290 (2010), pp. 890–893 (cit. on p. 90).

[Nag+13] Máté Nagy et al. “Context-dependent hierarchies in pigeons”. In: *Proceedings of the National Academy of Sciences* 110.32 (2013), pp. 13049–13054 (cit. on p. 23).

[PP01] Robert E Page Jr et al. “Aging and development in social insects with emphasis on the honey bee, Apis mellifera L.” In: *Experimental gerontology* 36.4-6 (2001), pp. 695–711 (cit. on p. 27).

[Pap+22] Marina Papadopoulou et al. “Emergence of splits and collective turns in pigeon flocks under predation”. In: *Royal Society open science* 9.2 (2022), p. 211898 (cit. on pp. 21, 23).

[PHH23] Marina Papadopoulou et al. “Diffusion during collective turns in bird flocks under predation”. In: *Frontiers in Ecology and Evolution* 11 (2023), p. 1198248 (cit. on p. 23).

[Pap+23] Marina Papadopoulou et al. “Dynamics of collective motion across time and species”. In: *Philosophical Transactions of the Royal Society B* 378.1874 (2023), p. 20220068 (cit. on pp. 38, 39, 86, 115).

[PHH97] Julia K Parrish et al. *Animal groups in three dimensions: how species aggregate*. Cambridge University Press, 1997 (cit. on p. 25).

[Par82] Brian L Partridge. “The structure and function of fish schools”. In: *Scientific american* 246.6 (1982), pp. 114–123 (cit. on pp. 33, 35).

[Pea+14a] Daniel J. G. Pearce et al. “Role of projection in the control of bird flocks”. In: *Proceedings of the National Academy of Sciences* 111.29 (2014), pp. 10422–10426. DOI: 10.1073/pnas.1402202111. URL: https://www.pnas.org/doi/abs/10.1073/pnas.1402202111 (cit. on p. 90).

[Pea+14b] Daniel JG Pearce et al. “Role of projection in the control of bird flocks”. In: *Proceedings of the National Academy of Sciences* 111.29 (2014), pp. 10422–10426 (cit. on pp. 65, 70).

[PRF10] Geoffrey Portelli et al. “Honeybees change their height to restore their optic flow”. In: *Journal of Comparative Physiology A* 196.4 (2010), pp. 307–313 (cit. on pp. 70, 90).

[Pot84] Wayne K Potts. “The chorus-line hypothesis of manoeuvre coordination in avian flocks”. In: *Nature* 309.5966 (1984), pp. 344–345 (cit. on p. 37).

[Pro+11] Andrea Procaccini et al. “Propagating waves in starling, Sturnus vulgaris, flocks under predation”. In: *Animal behaviour* 82.4 (2011), pp. 759–765 (cit. on p. 23).

[Qi+22] Jingtao Qi et al. “The emergence of collective obstacle avoidance based on a visual perception mechanism”. In: *Information Sciences* 582 (2022), pp. 850–864 (cit. on p. 70).

[Ren+16] Jiaping Ren et al. “Simulating flying insects using dynamics and data-driven noise modeling to generate diverse collective behaviors”. In: *PloS one* 11.5 (2016), e0155698 (cit. on p. 22).

[RGM10] Steven M Reppert et al. “Navigational mechanisms of migrating monarch butterflies”. In: *Trends in neurosciences* 33.9 (2010), pp. 399–406 (cit. on p. 31).

[Rey23] Andy M Reynolds. “Mosquito swarms shear harden”. In: *The European Physical Journal E* 46.12 (2023), p. 126 (cit. on pp. 27, 29).

[Rey87] Craig W Reynolds. “Flocks, herds and schools: A distributed behavioral model”. In: *Proceedings of the 14th annual conference on Computer graphics and interactive techniques*. 1987, pp. 25–34 (cit. on pp. 40, 41, 69, 84, 89).

[Ros+15] Sara Brin Rosenthal et al. “Revealing the hidden networks of interaction in mobile animal groups allows prediction of complex behavioral contagion”. In: *Proceedings of the National Academy of Sciences* 112.15 (2015), pp. 4690–4695 (cit. on p. 96).

[Rub+14] Michael Rubenstein et al. “Kilobot: A low cost robot with scalable operations designed for collective behaviors”. In: *Robotics and Autonomous Systems* 62.7 (2014), pp. 966–975 (cit. on p. 51).

[Ruf+03] Franck Ruffier et al. “Bio-inspired optical flow circuits for the visual guidance of micro air vehicles”. In: *Proceedings of the 2003 International Symposium on Circuits and Systems, 2003. ISCAS'03.* Vol. 3. IEEE. 2003, pp. III–III (cit. on p. 63).

[RF05] Franck Ruffier et al. “Optic flow regulation: the key to aircraft automatic guidance”. In: *Robotics and Autonomous Systems* 50.4 (2005), pp. 177–194 (cit. on pp. 70, 90).

[Rus+17] James C Russell et al. “Modeling collective animal movement through interactions in behavioral states”. In: *Journal of Agricultural, Biological and Environmental Statistics* 22 (2017), pp. 313–334 (cit. on p. 40).

[SP18] Takao Sasaki et al. “The psychology of superorganisms: collective decision making by insect societies”. In: *Annual Review of Entomology* 63.1 (2018), pp. 259–275 (cit. on p. 40).

[SSF22] Fabian Schilling et al. “On the Scalability of Vision-Based Drone Swarms in the Presence of Occlusions”. In: *IEEE Access* 10 (2022), pp. 28133–28146 (cit. on pp. 64, 70, 90).

[SP48] Theodore C Schneirla et al. “The army ant”. In: *Scientific American* 178.6 (1948), pp. 16–23 (cit. on pp. 31, 32).

[Ser+19] Julien R Serres et al. “Optic flow cues help explain altitude control over sea in freely flying gulls”. In: *Journal of the Royal Society Interface* 16.159 (2019), p. 20190486 (cit. on pp. 70, 90).

[SB14] Yilun Shang et al. “Influence of the number of topologically interacting neighbors on swarm dynamics”. In: *Scientific reports* 4.1 (2014), pp. 1–7 (cit. on pp. 50, 75, 90).

[Sha62] Evelyn Shaw. “The schooling of fishes”. In: *Scientific American* 206.6 (1962), pp. 128–141 (cit. on p. 33).

[Sha78] Evelyn Shaw. “Schooling fishes: the school, a truly egalitarian form of organization in which all members of the group are alike in influence, offers substantial benefits to its participants”. In: *American Scientist* 66.2 (1978), pp. 166–175 (cit. on p. 33).

[SW22] Yankai Shen et al. “Multi-UAV flocking control with individual properties inspired by bird behavior”. In: *Aerospace Science and Technology* 130 (2022), p. 107882 (cit. on p. 69).

[Shi+96] Naohiko Shimoyama et al. “Collective motion in a system of motile elements”. In: *Physical Review Letters* 76.20 (1996), p. 3870 (cit. on p. 42).

[Sil+13] Jesse L Silverberg et al. “Collective motion of humans in mosh and circle pits at heavy metal concerts”. In: *Physical review letters* 110.22 (2013), p. 228701 (cit. on p. 69).

[Sin77] Anthony Ronald Entrican Sinclair. “The African buffalo; a study of resource limitation of populations”. In: (1977) (cit. on pp. 21, 24).

[Sol+15] Alexandre P Solon et al. “Pattern formation in flocking models: A hydrodynamic description”. In: *Physical Review E* 92.6 (2015), p. 062111 (cit. on pp. 69, 90).

[SSF19] Enrica Soria et al. “The influence of limited visual sensing on the Reynolds flocking algorithm”. In: *2019 Third IEEE International Conference on Robotic Computing (IRC)*. IEEE. 2019, pp. 138–145 (cit. on pp. 65, 69, 90).

[SSF21] Enrica Soria et al. “Predictive control of aerial swarms in cluttered environments”. In: *Nature Machine Intelligence* 3.6 (2021), pp. 545–554 (cit. on p. 90).

[Str+13] Ariana Strandburg-Peshkin et al. “Visual sensory networks and effective information transfer in animal groups”. In: *Current Biology* 23.17 (2013), R709–R711 (cit. on pp. 69, 90).

[Str+76] Nicholas James Strausfeld et al. “Atlas of an insect brain”. In: (1976) (cit. on p. 55).

[Str11] Daniel Strömbom. “Collective motion from local attraction”. In: *Journal of theoretical biology* 283.1 (2011), pp. 145–151 (cit. on pp. 48, 86).

[Sum+08] David Sumpter et al. “Information transfer in moving animal groups”. In: *Theory in biosciences* 127.2 (2008), pp. 177–186 (cit. on p. 23).

[] *Supplemental Material Resources Of This Paper.*

[Swo05] Gregory A Sword. “Local population density and the activation of movement in migratory band-forming Mormon crickets”. In: *Animal Behaviour* 69.2 (2005), pp. 437–444 (cit. on p. 31).

[The+03] Guy Theraulaz et al. “The formation of spatial patterns in social insects: from simple behaviours to complex structures”. In: *Philosophical Transactions of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences* 361.1807 (2003), pp. 1263–1282 (cit. on p. 21).

[Ull81] Shimon Ullman. “Analysis of visual motion by biological and computer systems”. In: *Computer* 14.08 (1981), pp. 57–69 (cit. on pp. 61, 73).

[Ver+23] Matthieu Verdoucq et al. “Bio-inspired 3D flocking algorithm with minimal information transfer for drones swarms”. In: *2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE. 2023, pp. 8833–8838 (cit. on p. 53).

[Vic+95] Tamás Vicsek et al. “Novel type of phase transition in a system of self-driven particles”. In: *Physical review letters* 75.6 (1995), p. 1226 (cit. on pp. 21, 26, 30, 41, 42, 69, 90).

[VZ12] Tamás Vicsek et al. “Collective motion”. In: *Physics reports* 517.3-4 (2012), pp. 71–140 (cit. on pp. 20, 21, 69, 72, 90).

[Vir+14] Csaba Virágh et al. “Flocking algorithm for autonomous flying robots”. In: *Bioinspiration & biomimetics* 9.2 (2014), p. 025012 (cit. on pp. 69, 90).

[VB98] Han de Vries et al. “Modelling collective foraging by means of individual behaviour rules in honey-bees”. In: *Behavioral Ecology and Sociobiology* 44 (1998), pp. 109–124 (cit. on p. 31).

[WL19] Xiong Wang et al. “Collective behaviors through social interactions in bird flocks”. In: *IEEE Circuits and Systems Magazine* 19.3 (2019), pp. 6–22 (cit. on p. 69).

[Wei+01] Henri Weimerskirch et al. “Energy saving in flight formation”. In: *Nature* 413.6857 (2001), pp. 697–698 (cit. on p. 36).

[Wen+12] Guanghui Wen et al. “Flocking of multi-agent dynamical systems with intermittent nonlinear velocity measurements”. In: *International Journal of Robust and Nonlinear Control* 22.16 (2012), pp. 1790–1805 (cit. on p. 90).

[Wey+08] Tina Wey et al. “Social network analysis of animal behaviour: a promising tool for the study of sociality”. In: *Animal behaviour* 75.2 (2008), pp. 333–344 (cit. on p. 23).

[Yam+05] Yoshihiro Yamazaki et al. “Periodic growth of bacterial colonies”. In: *Physica D: Nonlinear Phenomena* 205.1-4 (2005), pp. 136–153 (cit. on p. 27).

[Yat+09] Christian A Yates et al. “Inherent noise can facilitate coherence in collective swarm motion”. In: *Proceedings of the National Academy of Sciences* 106.14 (2009), pp. 5464–5469 (cit. on p. 30).

[ZSS17] Ayham Zaitouny et al. “Modelling and tracking the flight dynamics of flocking pigeons based on real GPS data (small flock)”. In: *Ecological Modelling* 344 (2017), pp. 62–72 (cit. on p. 26).

[ZZC10] Hai-Tao Zhang et al. “A general alignment repulsion algorithm for flocking of multi-agent systems”. In: *IEEE Transactions on Automatic Control* 56.2 (2010), pp. 430–435 (cit. on p. 90).

[Zha+10] He-Peng Zhang et al. “Collective motion and density fluctuations in bacterial colonies”. In: *Proceedings of the National Academy of Sciences* 107.31 (2010), pp. 13626–13630 (cit. on p. 21).

[Zie+18] Adam K Zienkiewicz et al. “Data-driven modelling of social forces and collective behaviour in zebrafish”. In: *Journal of Theoretical Biology* 443 (2018), pp. 39–51 (cit. on p. 86).

ANNEXES

A. Supplementary information (Extra Results): Modeling collective behaviors from optic flow and retinal cues.

A.1. Extra Results

]

A.1.1. Visual Model

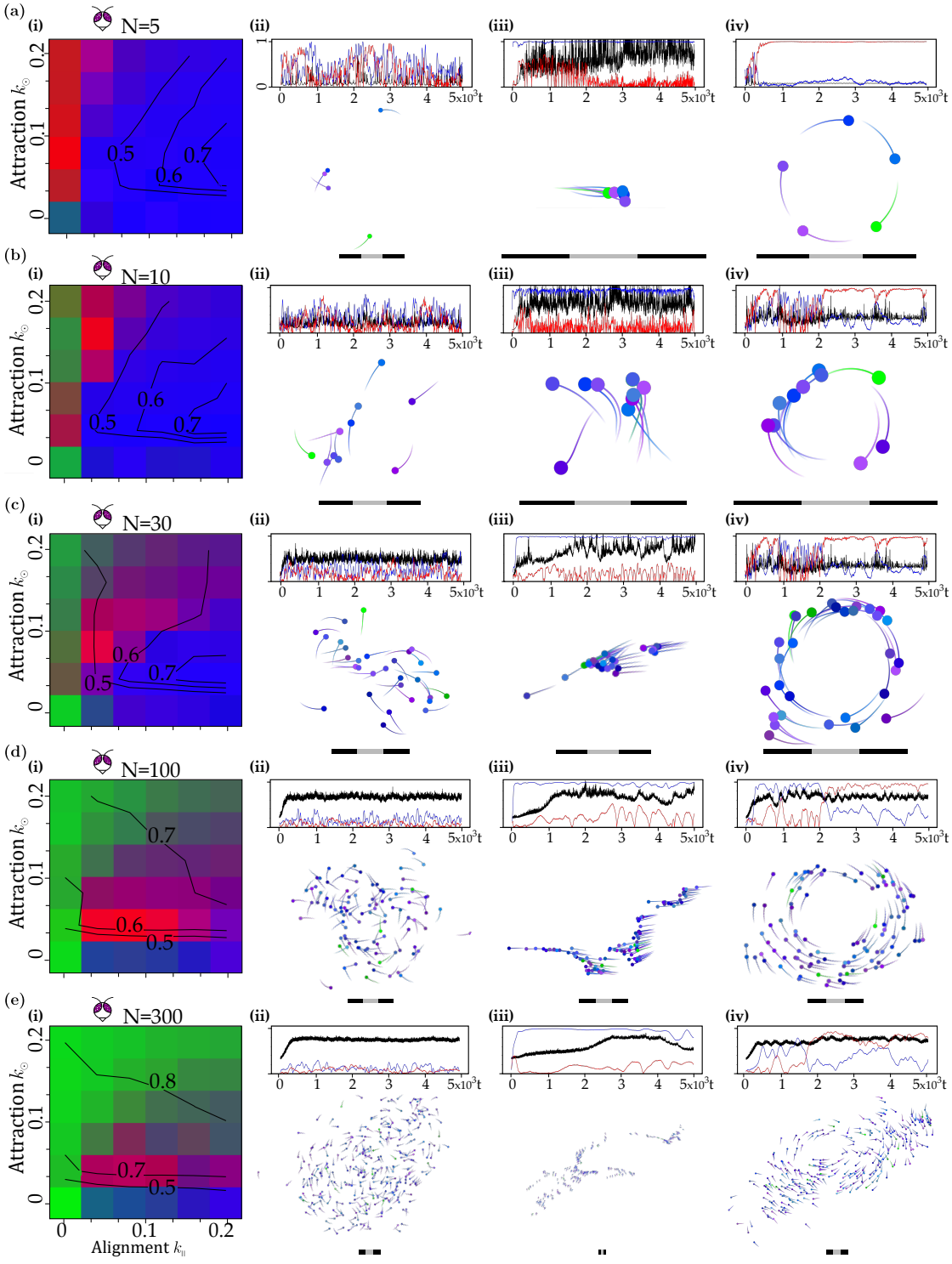

Figure A.2.: Phase Diagram and Phase Reproduction for different Flock sizes for the Visual model. i) Phase Diagram, ii)Swarming, iii)Schooling, iv)Milling. a)N=5, b)N=10, c)N=30, d)N=100, e)N=300,

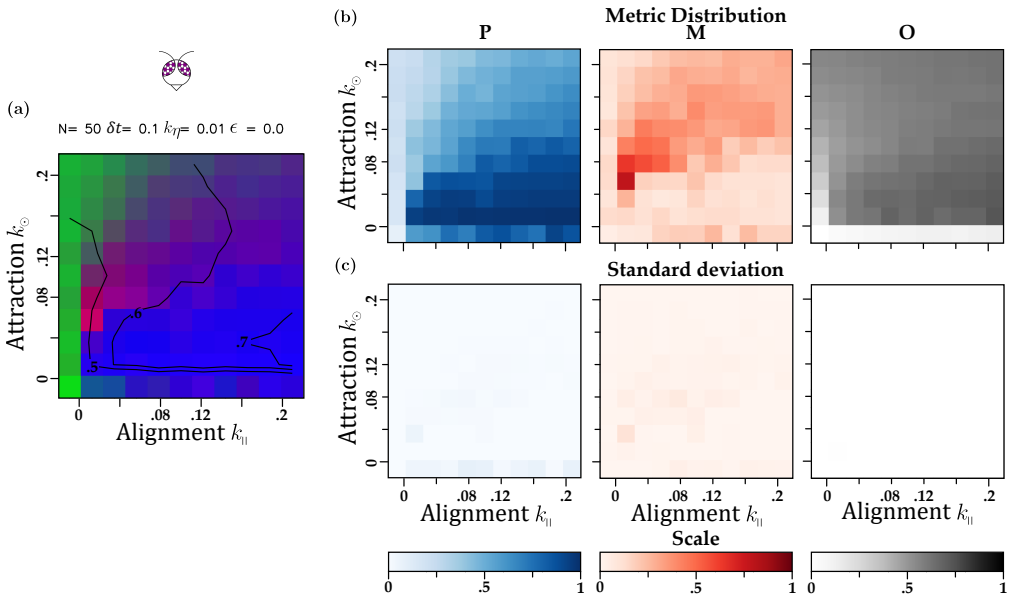

Figure A.3.: Phase diagram for the Visual model of a flock size of 50 individuals, k_η =0.01 and no blind angle. a) Integrated metrics, b) Separated metrics and c) Variance between the 10 experiments final metrics..

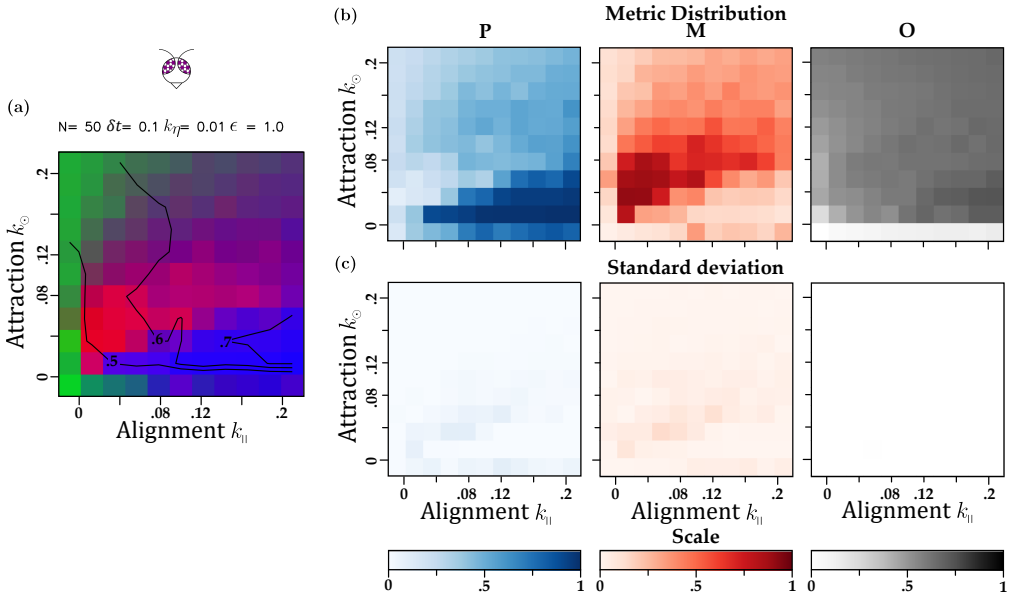

Figure A.4.: Phase diagram for the Visual model of a flock size of 50 individuals, k_η =0.01 and maximum blind angle. a) Integrated metrics, b) Separated metrics and c) Variance between the 10 experiments final metrics..

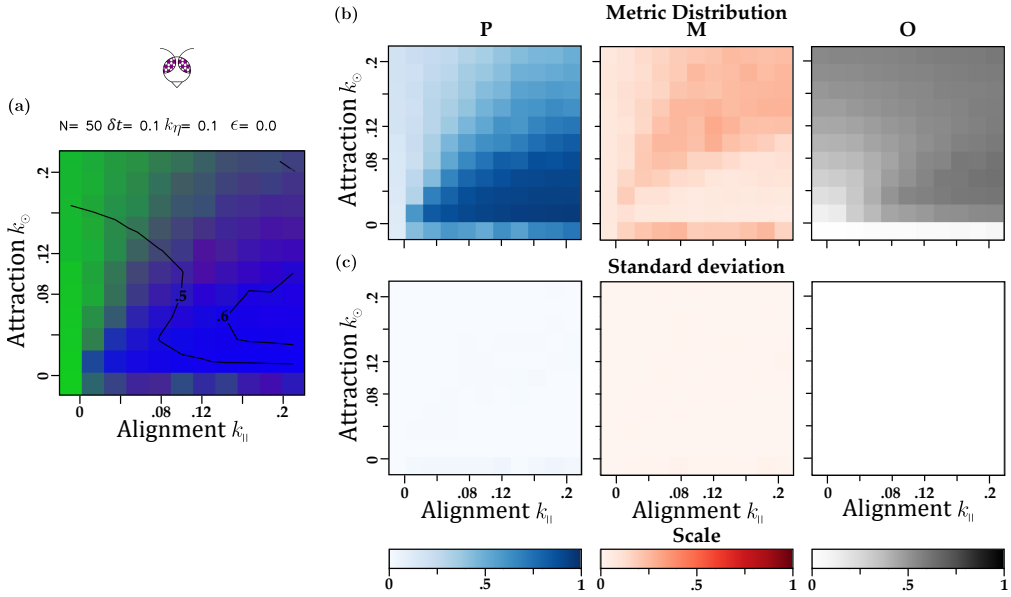

Figure A.5.: Phase diagram for the Visual model of a flock size of 50 individuals, k_η =0.1 and no blind angle. a) Integrated metrics, b) Separated metrics and c) Variance between the 10 experiments final metrics..

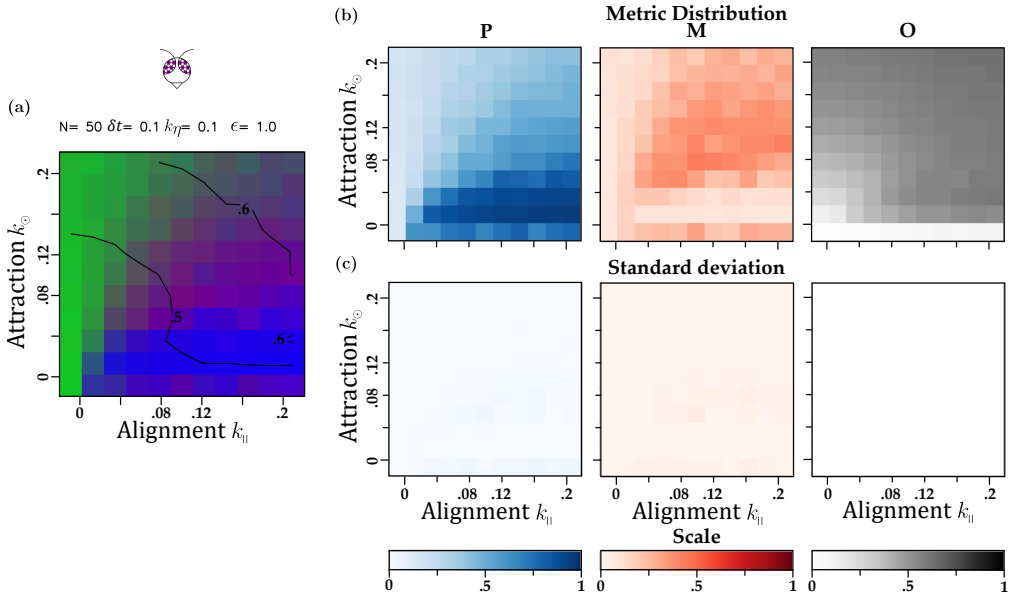

Figure A.6.: Phase diagram for the Visual model of a flock size of 50 individuals, k_η =0.1 and maximum blind angle. a) Integrated metrics, b) Separated metrics and c) Variance between the 10 experiments final metrics..

A.1.2. Omniscient Model

A.2. Video Snapshots

A.2.1. Time Step Analysis

A.2.1.1. Visual

A.2.1.2. Omniscient

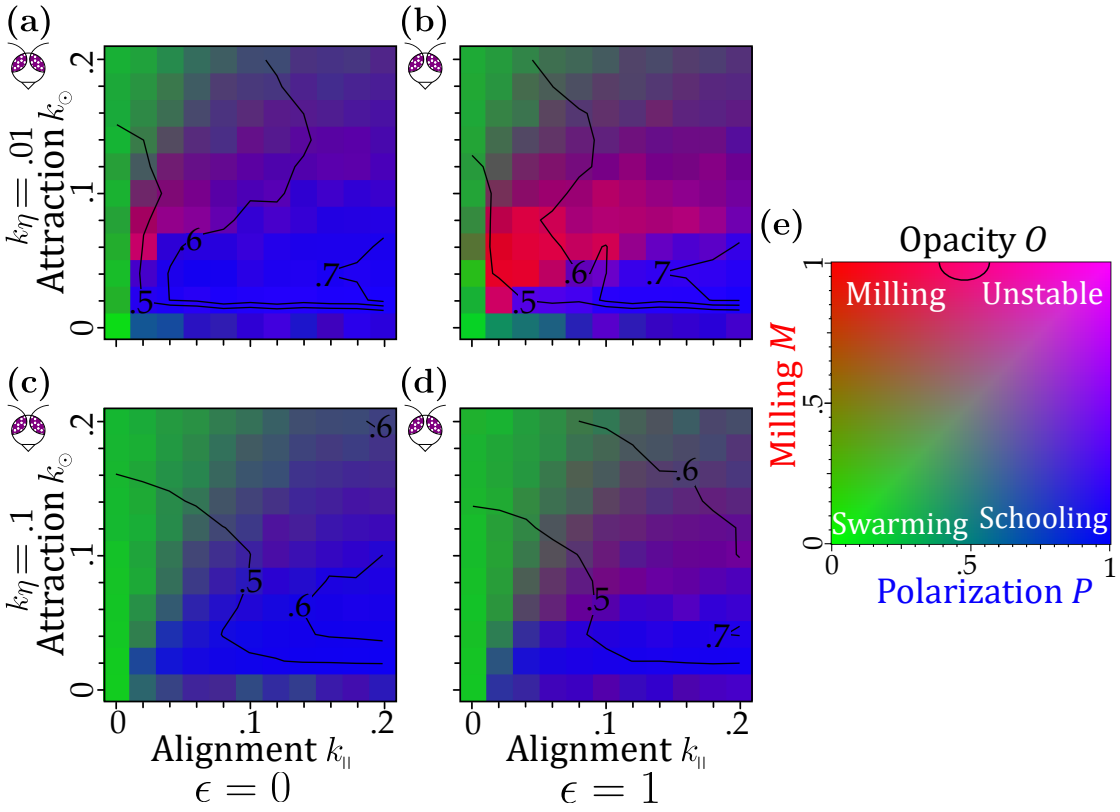

Figure A.1.: Phase diagrams for $N = 50$ individuals. These phase diagrams are for: (a)-(c) noise strength $k_\eta = 0.01$, where there is no blind angle ($\epsilon = 0$) and maximum ($\epsilon = 1$), respectively. (b)-(d) noise strength $k_\eta = 0.1$, where there is no blind angle ($\epsilon = 0$) and maximum ($\epsilon = 1$), respectively In (e), is illustrated how, in the phase space $(k_\parallel, k_\odot)$, the colors represents different values of P and M, and contours show the opacity O; these colors also represents the phases discussed earlier.

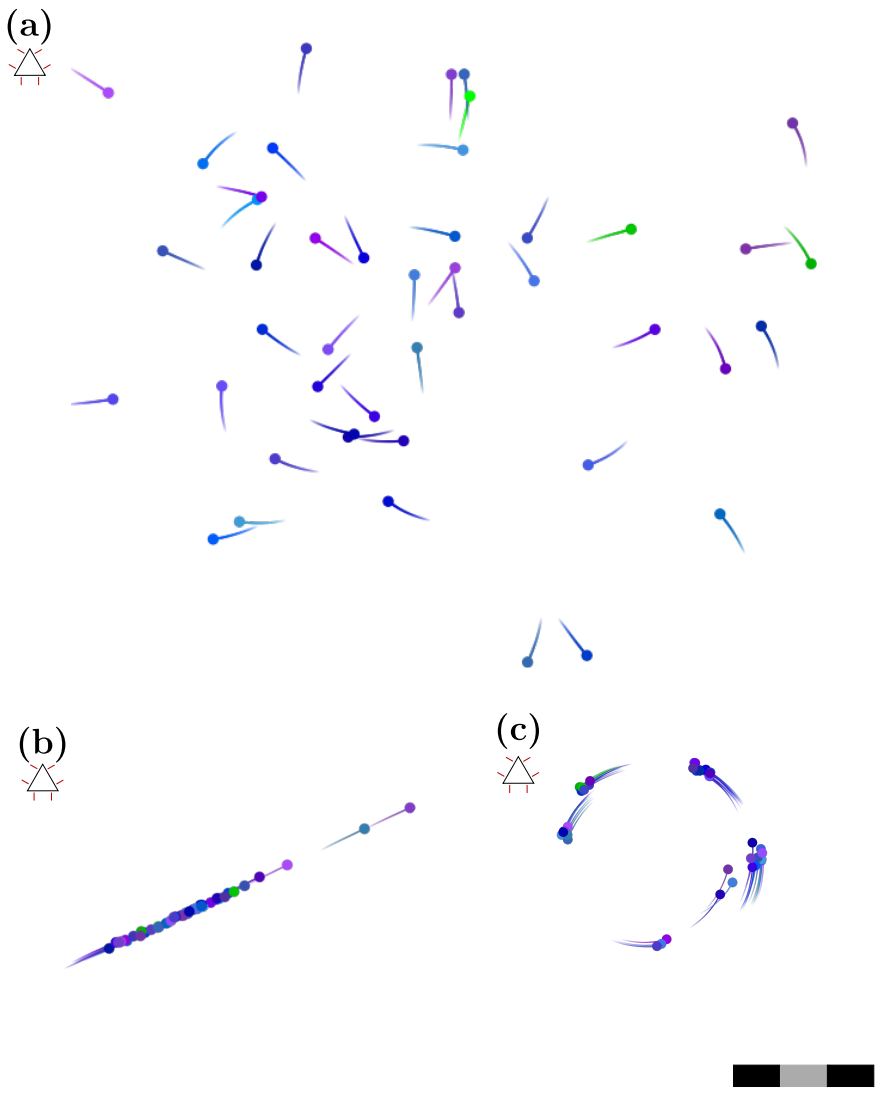

Figure A.7.: Illustration of the phases observed for the omniscient model and a flock size of $N = 50$ individuals. With $k_\eta = 0.01$. The scale bar measures $30a$ ($10a$ each section). We observe three phases: (a) Swarming ($k_\odot = 0.06$, $k_\parallel = 0, \epsilon = 0$); (b) Schooling ($k_\odot = 0.06$, $k_\parallel = 0.2, \epsilon = 1$); (c)Milling ($k_\odot = 0.06$, $k_\parallel = 0.06, \epsilon = 1$).

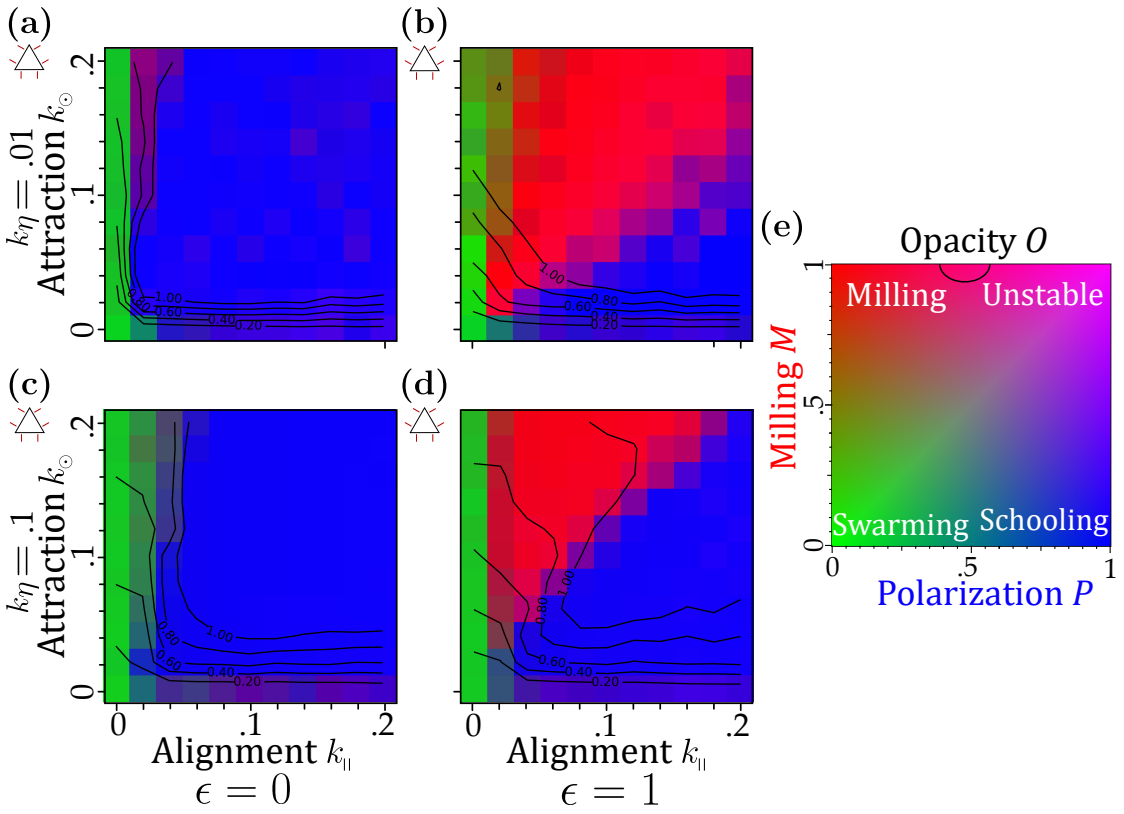

Figure A.8.: Phase diagrams for omniscient model and flock size of $N = 50$ individuals. These phase diagrams are for: (a)-(c) noise strength $k_\eta = 0.01, \epsilon = 0$ and $\epsilon = 1$, respectively. (b)-(d) noise strength $k_\eta = 0.1, \epsilon = 0$ and $\epsilon = 1$) respectively In (e), is illustrated how, in the phase space $(k_\parallel, k_\odot)$, the colors represents different values of P and M, and contours show the opacity O; these colors also represents the phases discussed earlier.

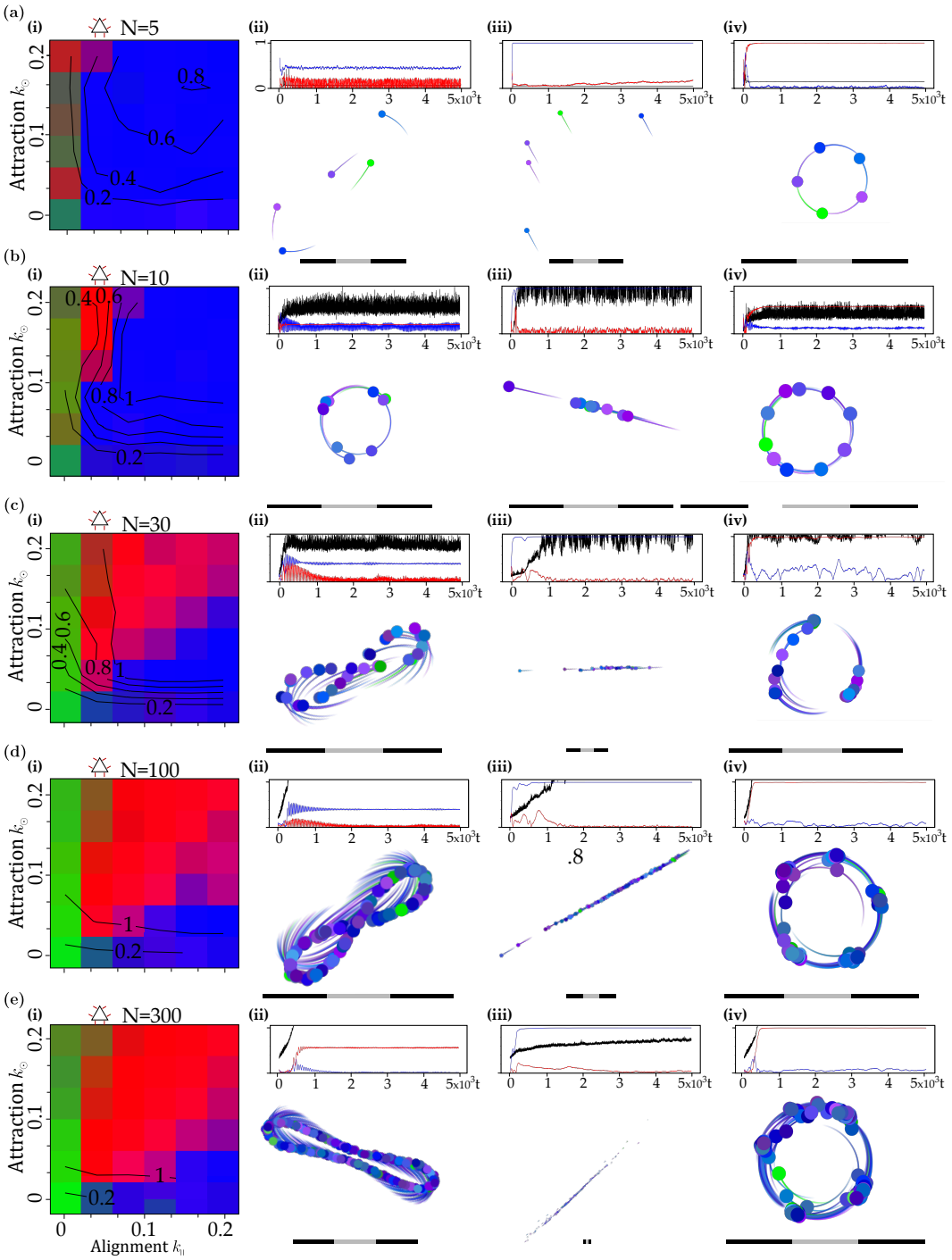

Figure A.9.: Phase Diagram and Phase Reproduction for different Flock sizes for the Visual model. i) Phase Diagram, ii)Swarming, iii)Schooling, iv)Milling. a)N=5, b)N=10, c)N=30, d)N=100, e)N=300.

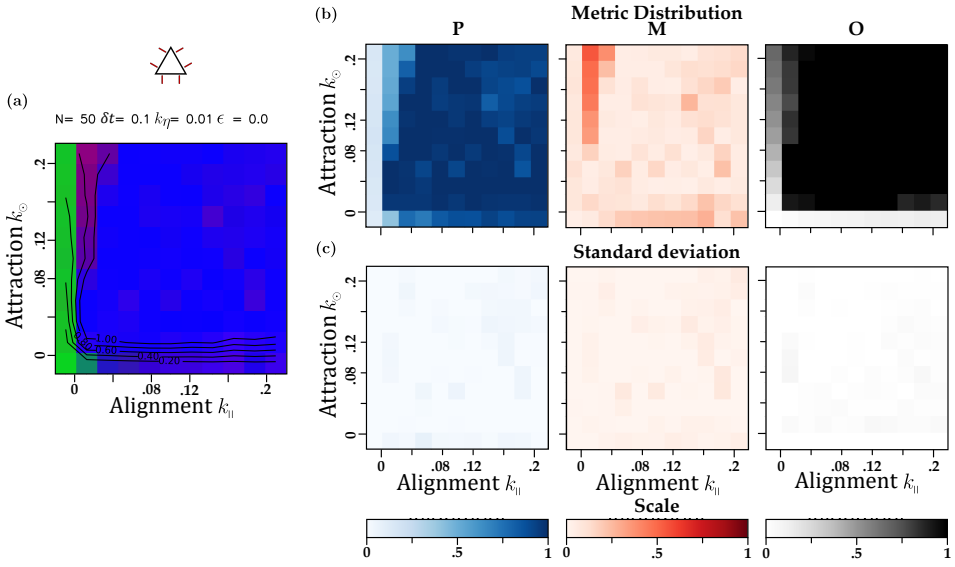

Figure A.10.: Phase diagram for the Omniscient model of a flock size of 50 individuals, k_η =0.01 and no blind angle. a) Integrated metrics, b) Separated metrics and c) Variance between the 10 experiments final metrics..

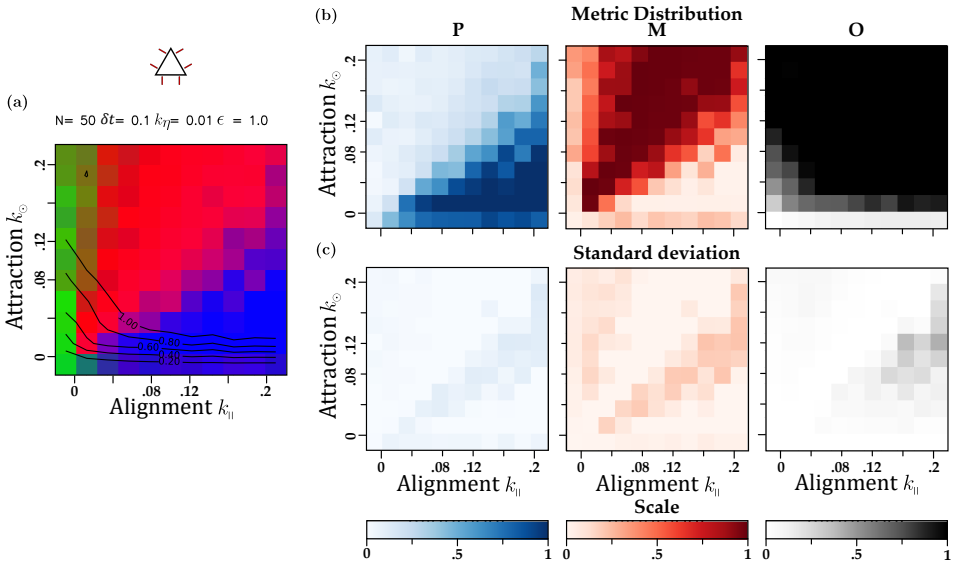

Figure A.11.: Phase diagram for the Omniscient model of a flock size of 50 individuals, k_η =0.01 and maximum blind angle. a) Integrated metrics, b) Separated metrics and c) Variance between the 10 experiments final metrics..

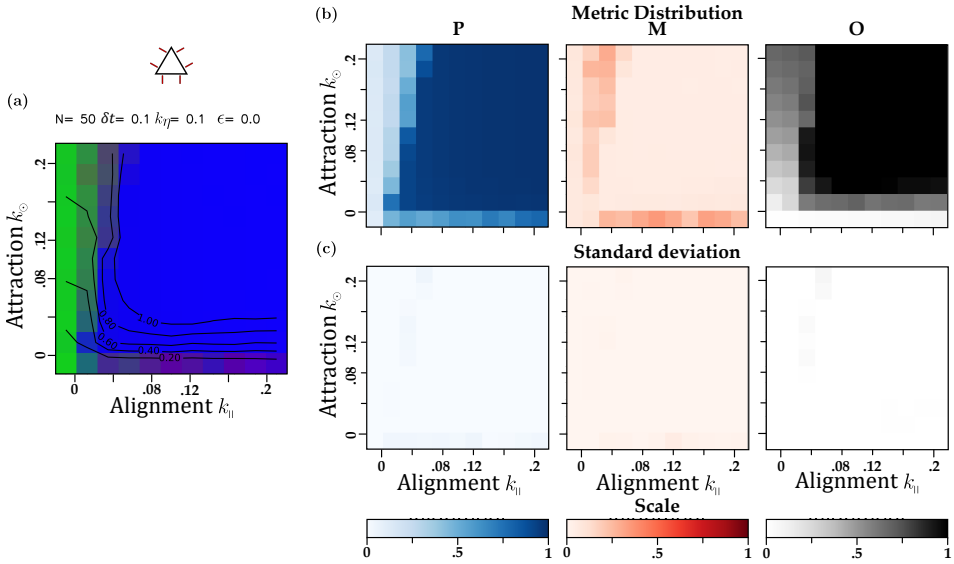

Figure A.12.: Phase diagram for the Omniscient model of a flock size of 50 individuals, k_η =0.1 and no blind angle. a) Integrated metrics, b) Separated metrics and c) Variance between the 10 experiments final metrics..

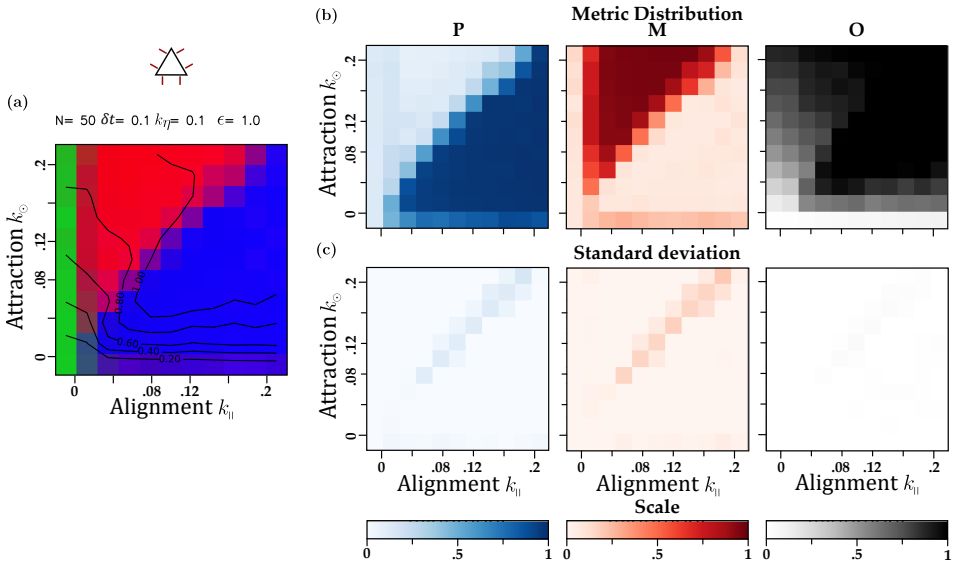

Figure A.13.: Phase diagram for the Omniscient model of a flock size of 50 individuals, k_η =0.1 and maximum blind angle. a) Integrated metrics, b) Separated metrics and c) Variance between the 10 experiments final metrics..

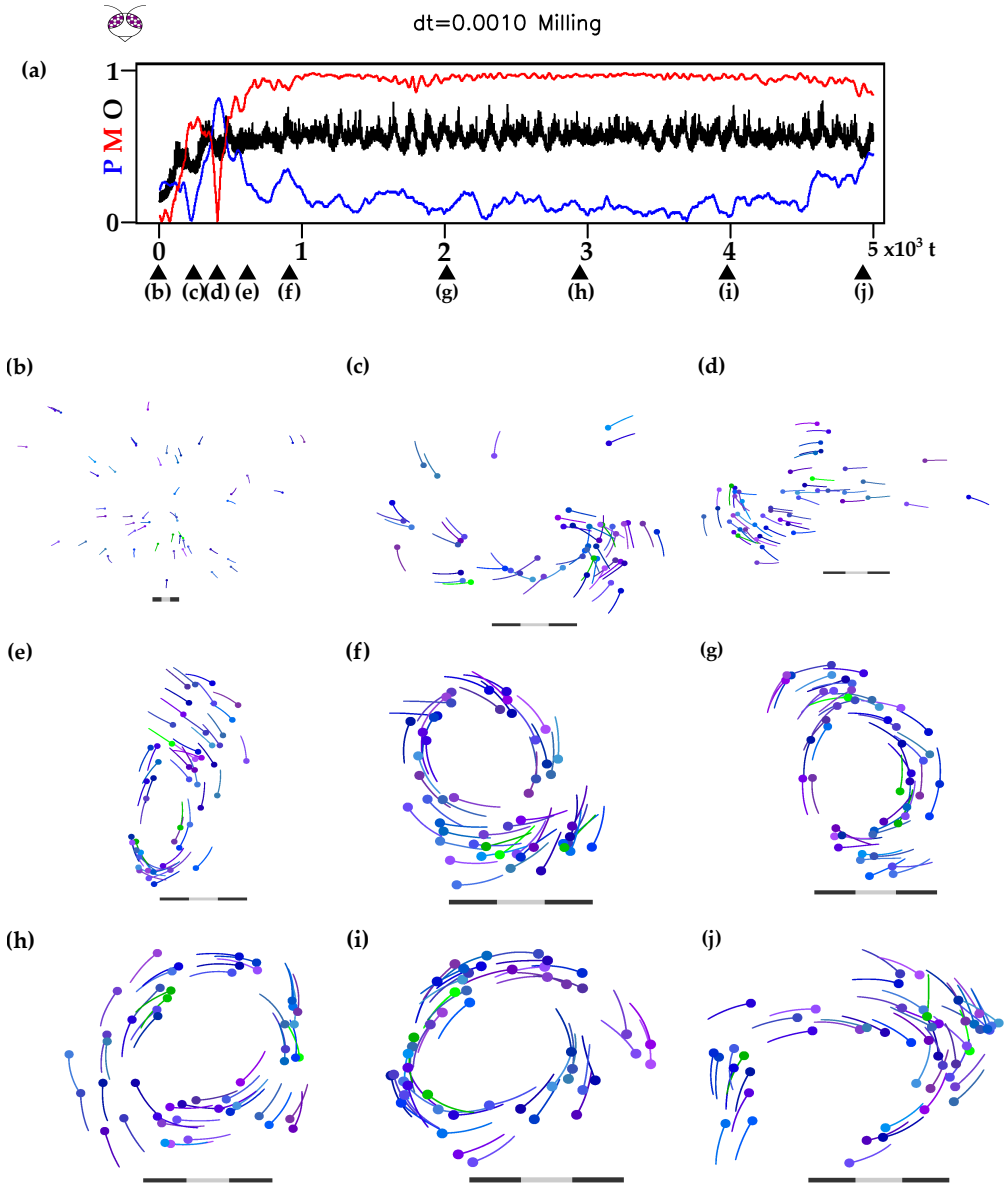

Figure A.14.: Snapshot of the time step stability at δt=0.001 for the Visual model. Reproduction of a milling phase for a flock size of 50 individuals, $k_{\odot}$ =0.06,$k_{\parallel}$ =0.06 and k_{η} =0.01. a) Evolution of the flock metrics, b-j different snapshots. .

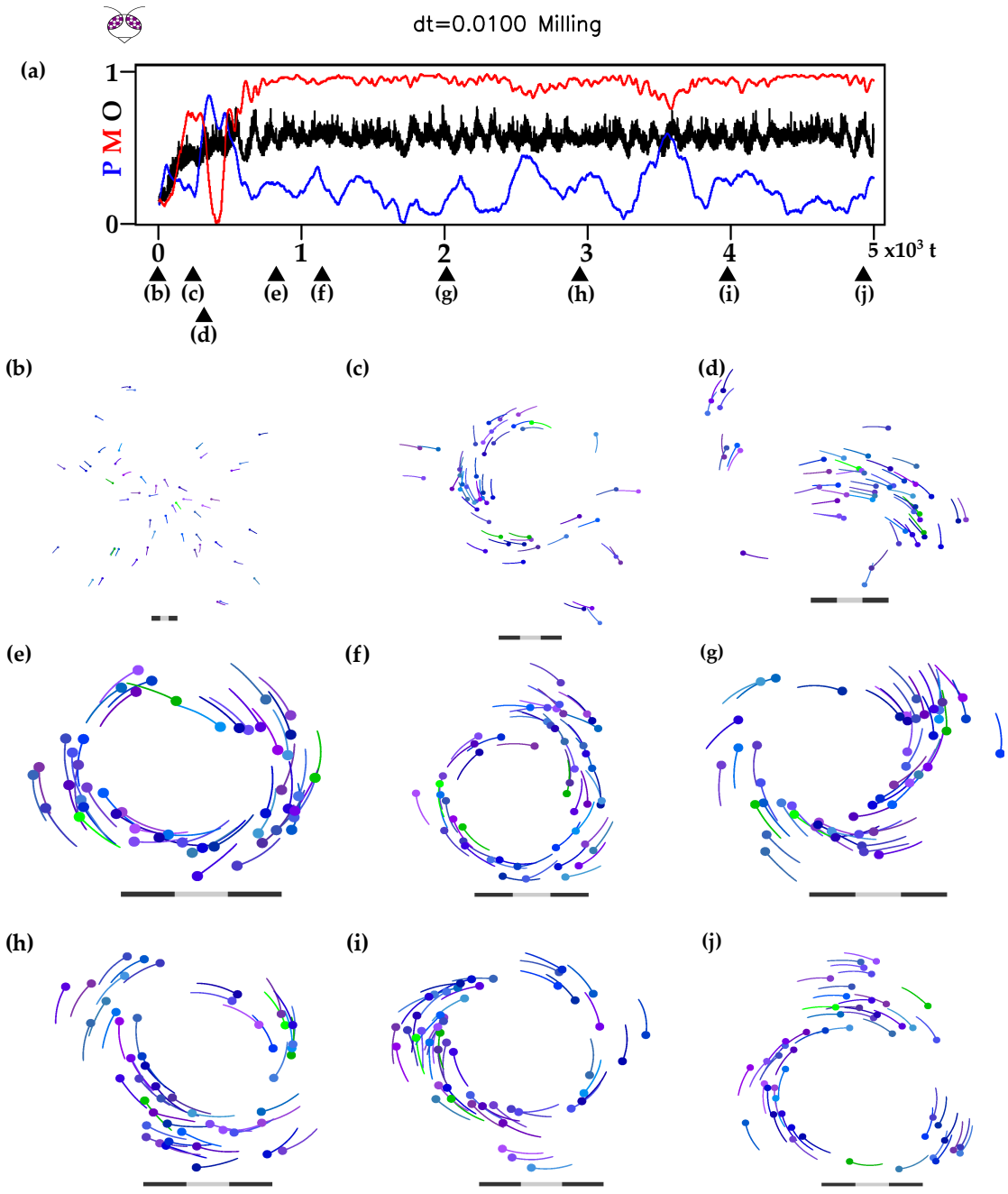

Figure A.15.: Snapshot of the time step stability at δt=0.01 for the Visual model. Reproduction of a milling phase for a flock size of 50 individuals, $k_{\odot}$ =0.06,$k_{\parallel}$ =0.06 and k_{η} =0.01. a) Evolution of the flock metrics, b-j different snapshots. .

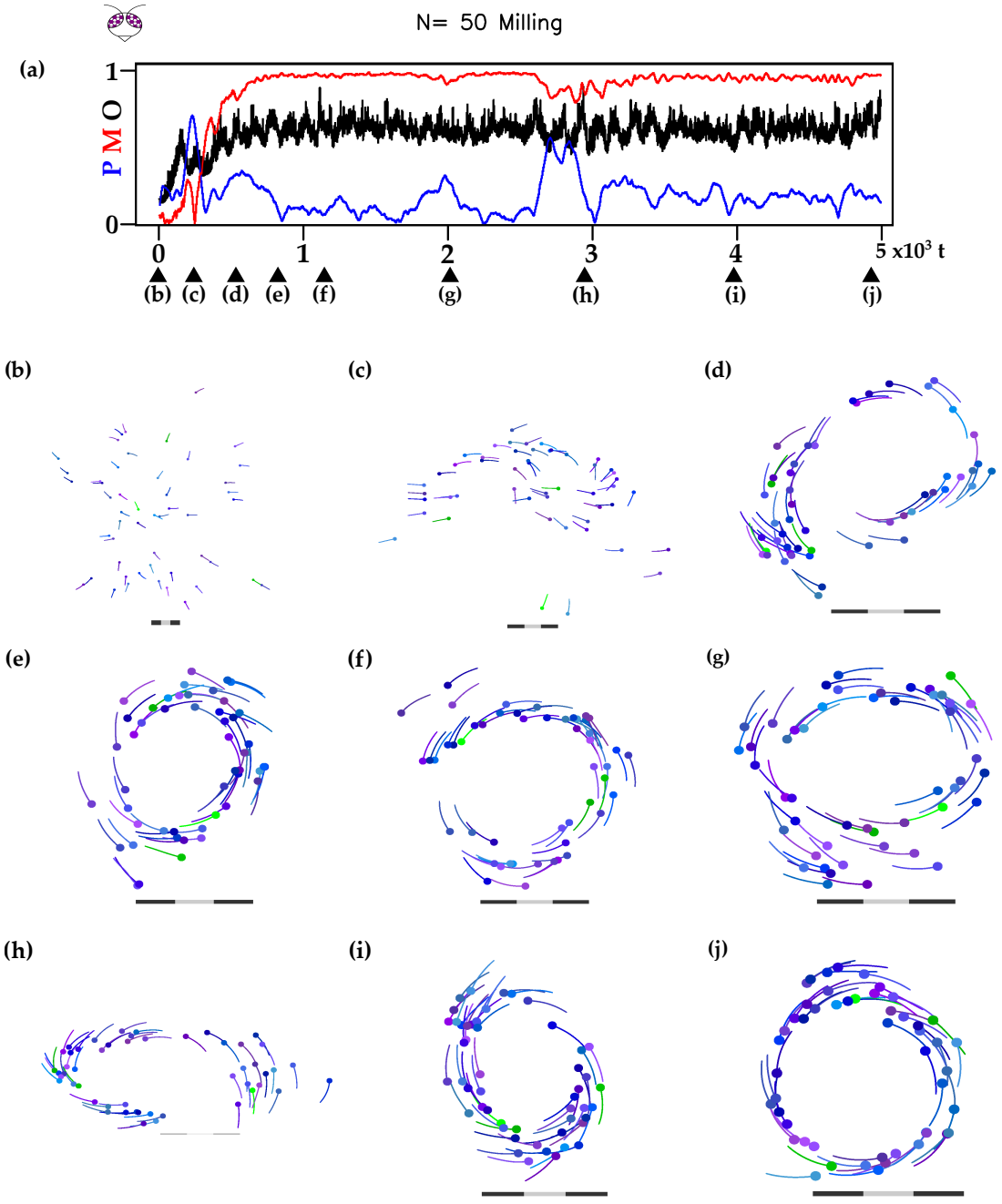

Figure A.16.: Snapshot of the time step stability at δt=0.1 for the Visual model. Reproduction of a milling phase for a flock size of 50 individuals, $k_{\odot}$ =0.06,$k_{\parallel}$ =0.06 and k_{η} =0.01. a) Evolution of the flock metrics, b-j different snapshots. .

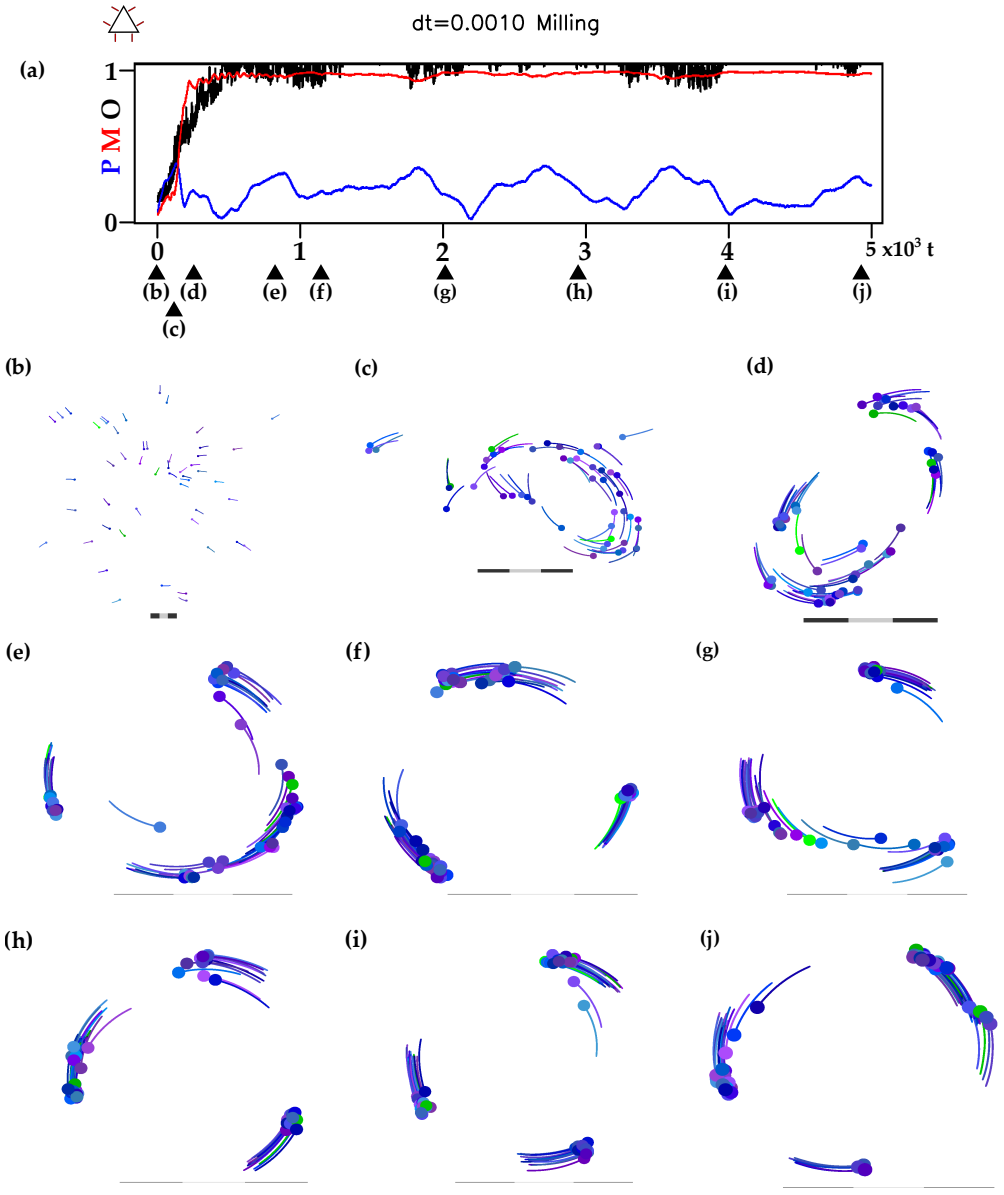

Figure A.17.: Snapshot of the time step stability at δt=0.001 for the Omniscient model. Reproduction of a milling phase for a flock size of 50 individuals, $k_{\odot}$ =0.06,$k_{\parallel}$ =0.06 and k_{η} =0.01. a) Evolution of the flock metrics, b-j different snapshots. .

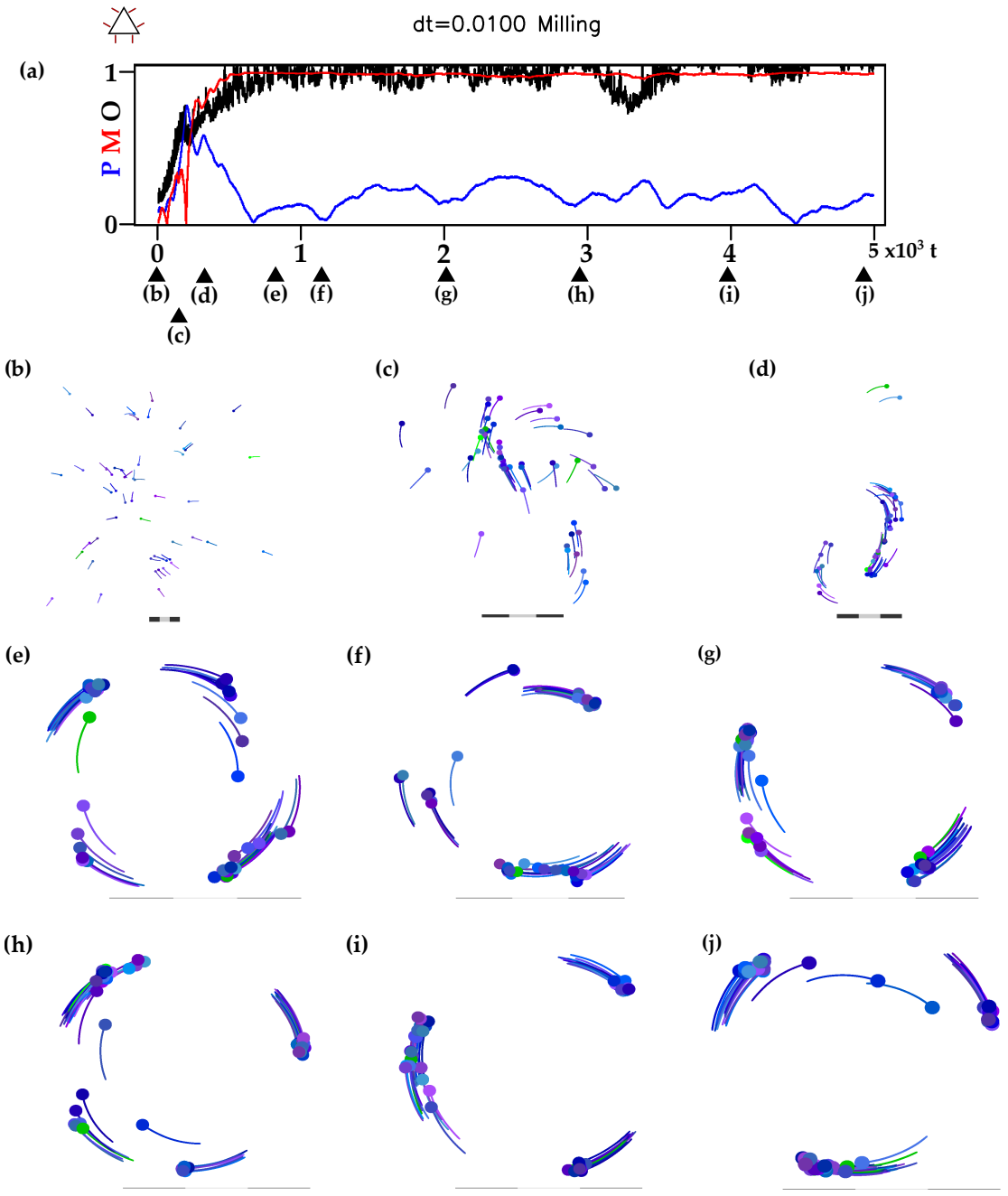

Figure A.18.: Snapshot of the time step stability at δt=0.01 for the Omniscient model. Reproduction of a milling phase for a flock size of 50 individuals, $k_{\odot}$ =0.06,$k_{\parallel}$ =0.06 and k_{η} =0.01. a) Evolution of the flock metrics, b-j different snapshots.

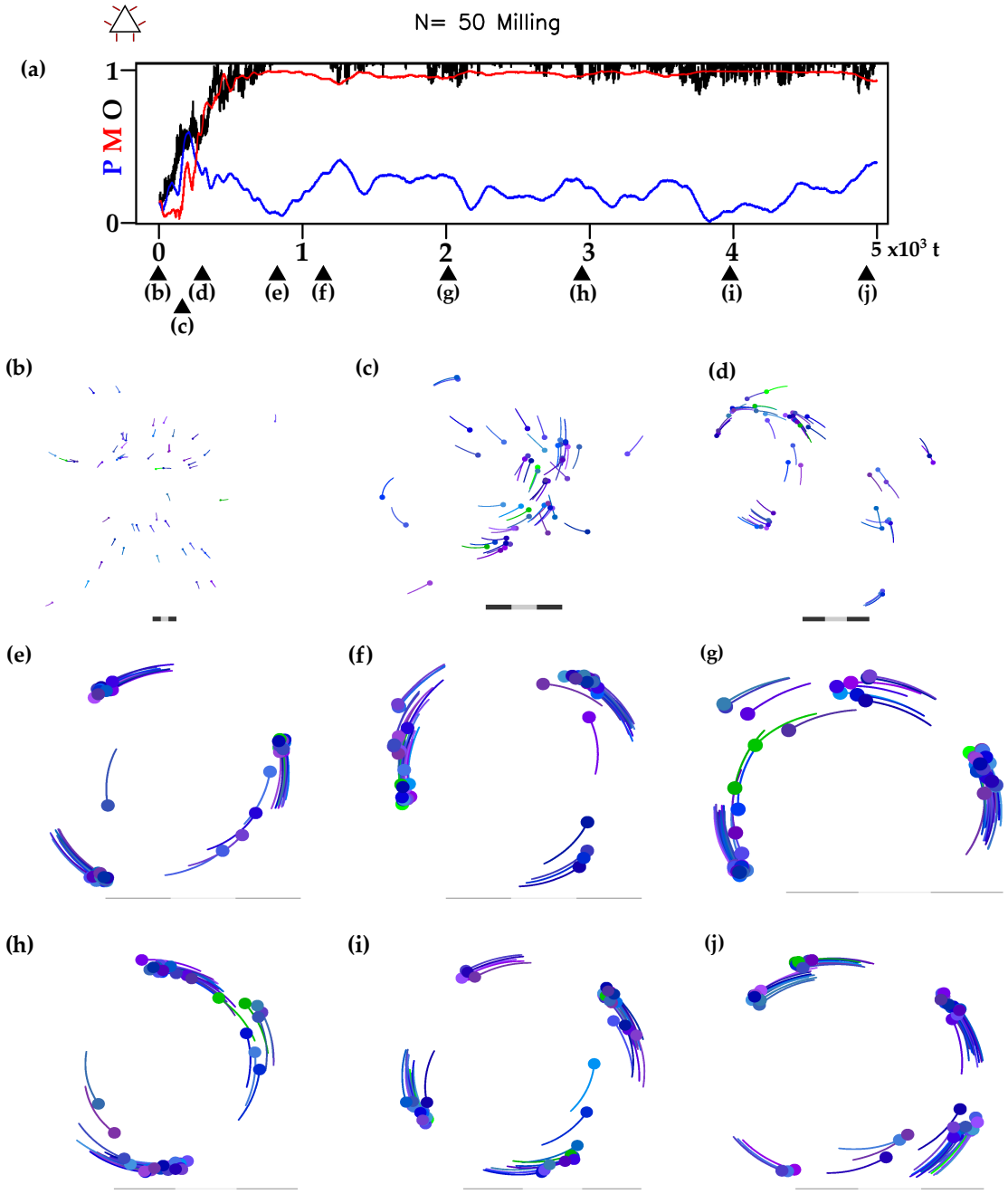

Figure A.19.: Snapshot of the time step stability at δt=0.1 for the Omniscient model. Reproduction of a milling phase for a flock size of 50 individuals, $k_{\odot}$ =0.06,$k_{\parallel}$ =0.06 and k_{η} =0.01. a) Evolution of the flock metrics, b-j different snapshots. .

A.3. Phases Reproduction $k_\eta = 0.01$ Flock Size=50

A.3.1. Visual

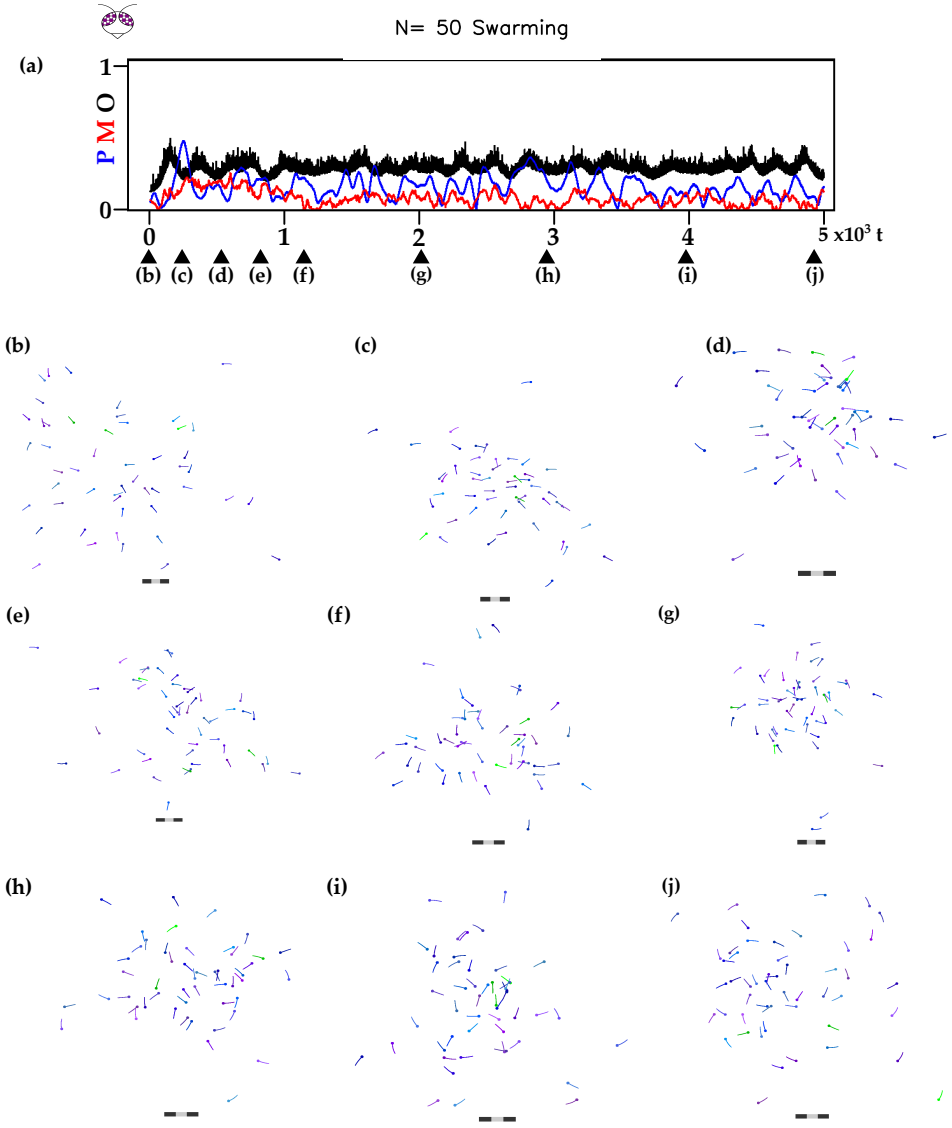

Figure A.20.: Swarming phase reproduction for a flock size of 50 individuals for the Visual model. k_η =0.06 k_η =0.0 k_η =0.01 and no blind angle .

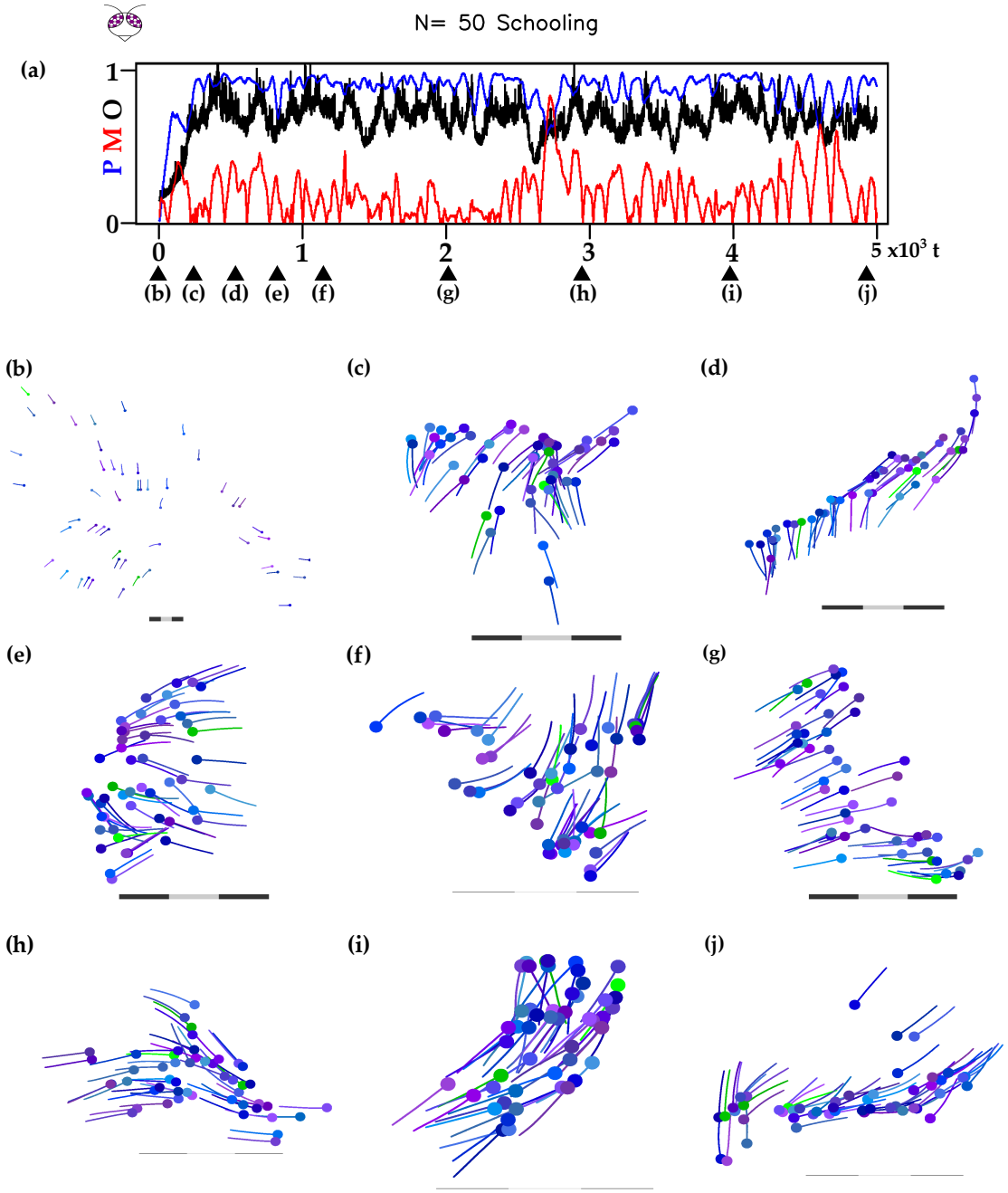

Figure A.21.: Schooling phase reproduction for a flock size of 50 individuals for the Visual model. k_η =0.06 k_η =0.2 k_η =0.01 and maximum blind angle .

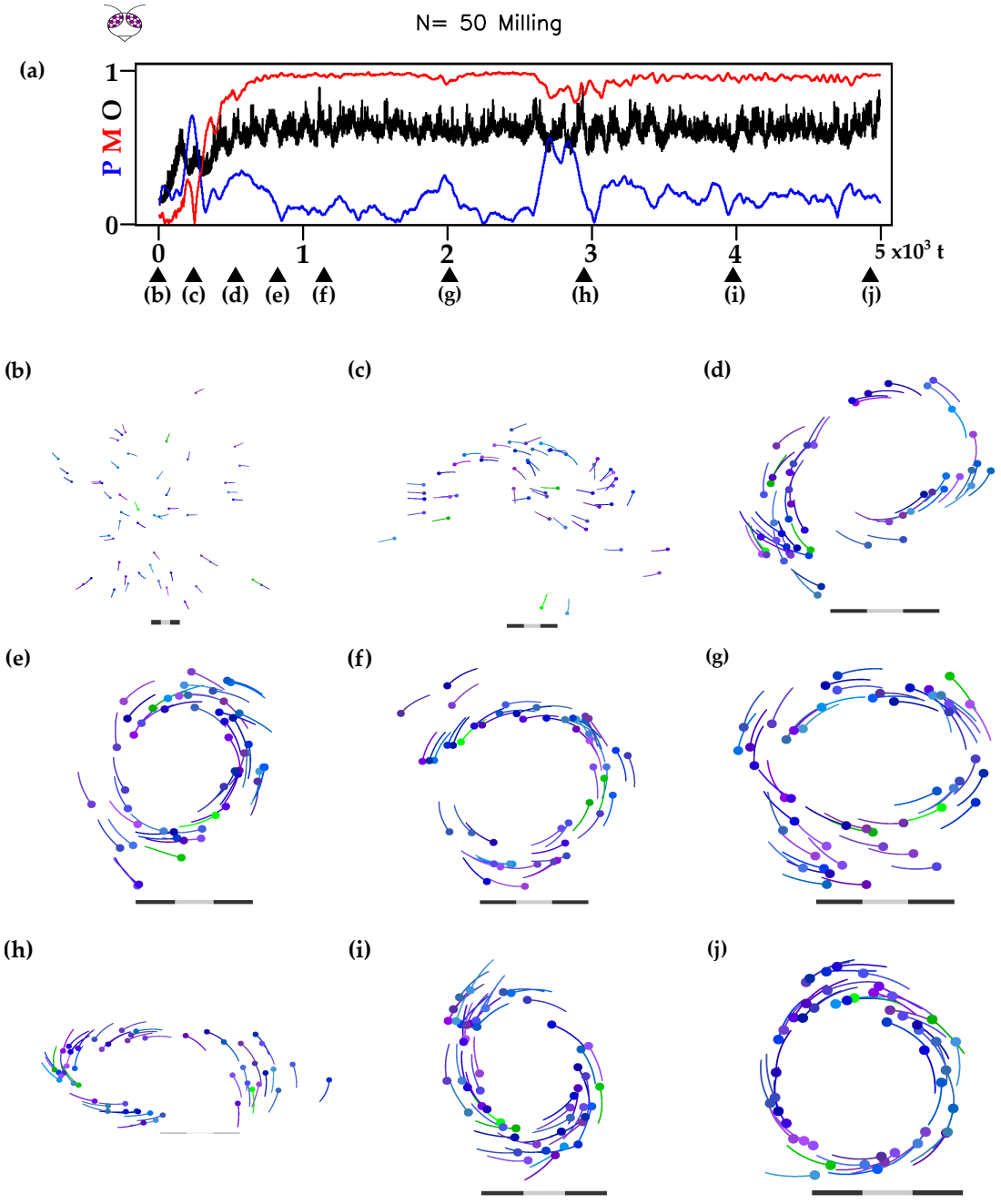

Figure A.22.: Milling phase reproduction for a flock size of 50 individuals for the Visual model. k_η =0.06 k_η =0.06 k_η =0.01 and maximum blind angle .

A.3.2. Omniscient

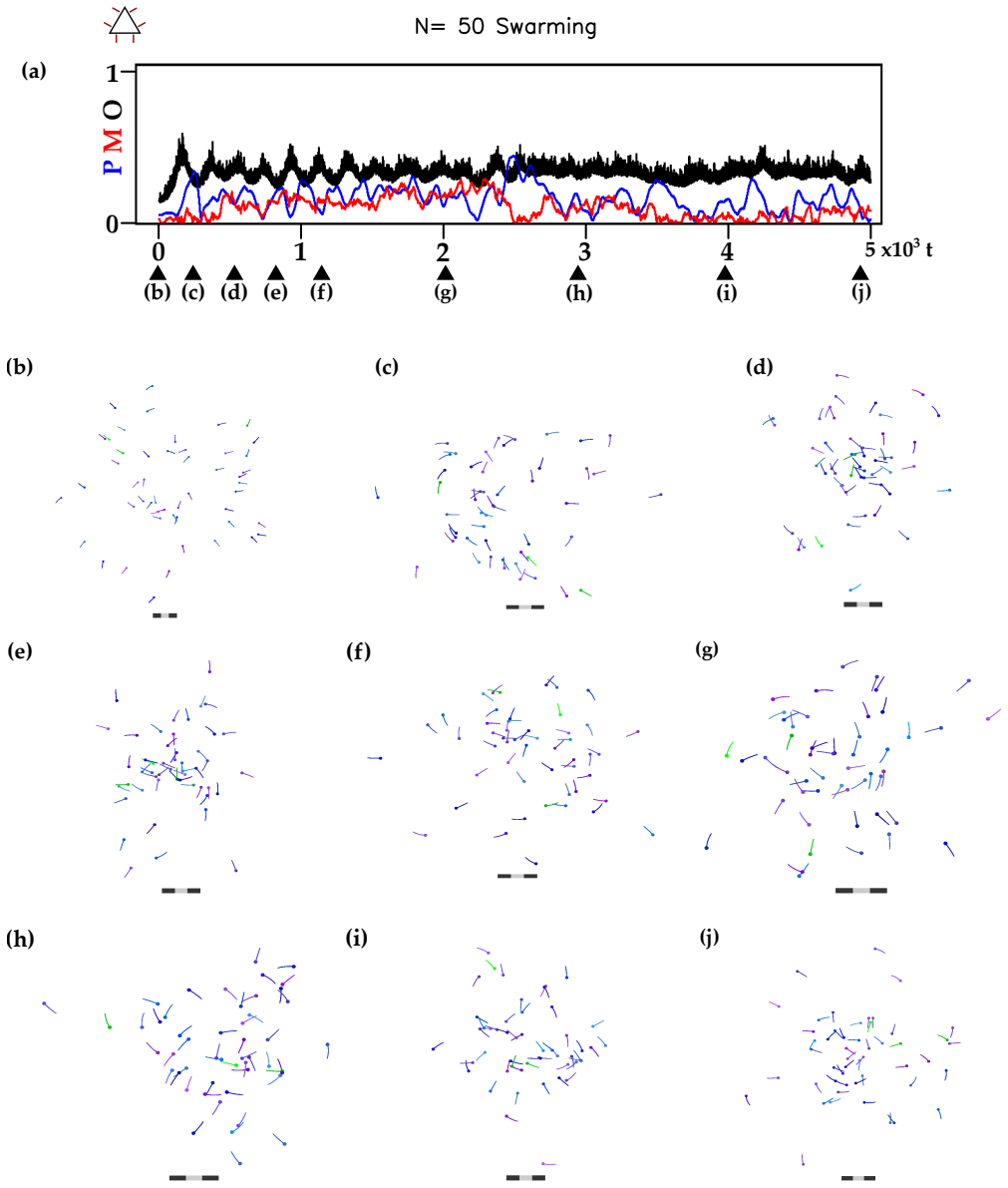

Figure A.23.: Swarming phase reproduction for a flock size of 50 individuals for the Omniscient model. k_η =0.06 k_η =0.0 k_η =0.01 and no blind angle .

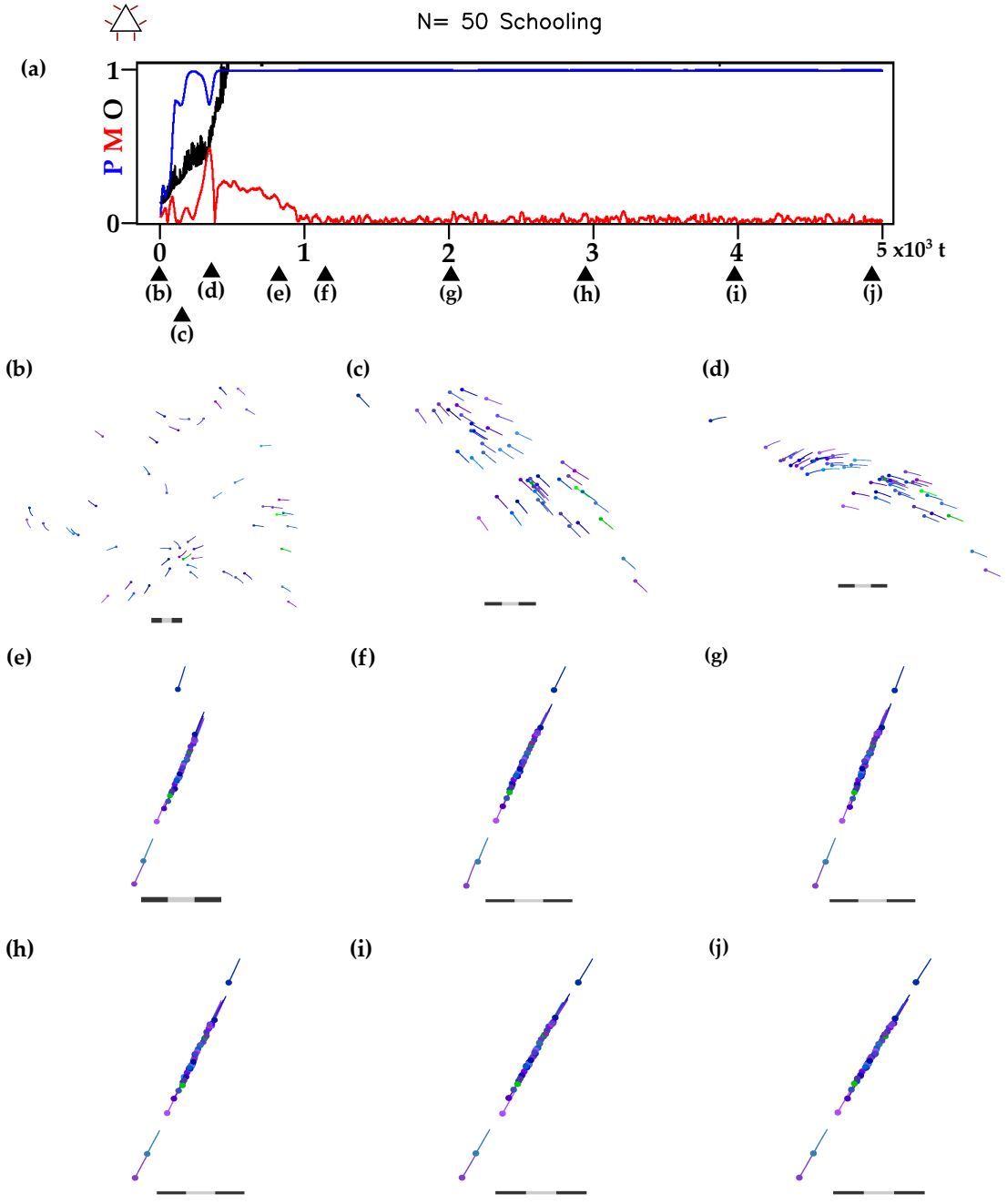

Figure A.24.: Schooling phase reproduction for a flock size of 50 individuals for the Omniscient model. k_η =0.06 k_η =0.2 k_η =0.01 and maximum blind angle .

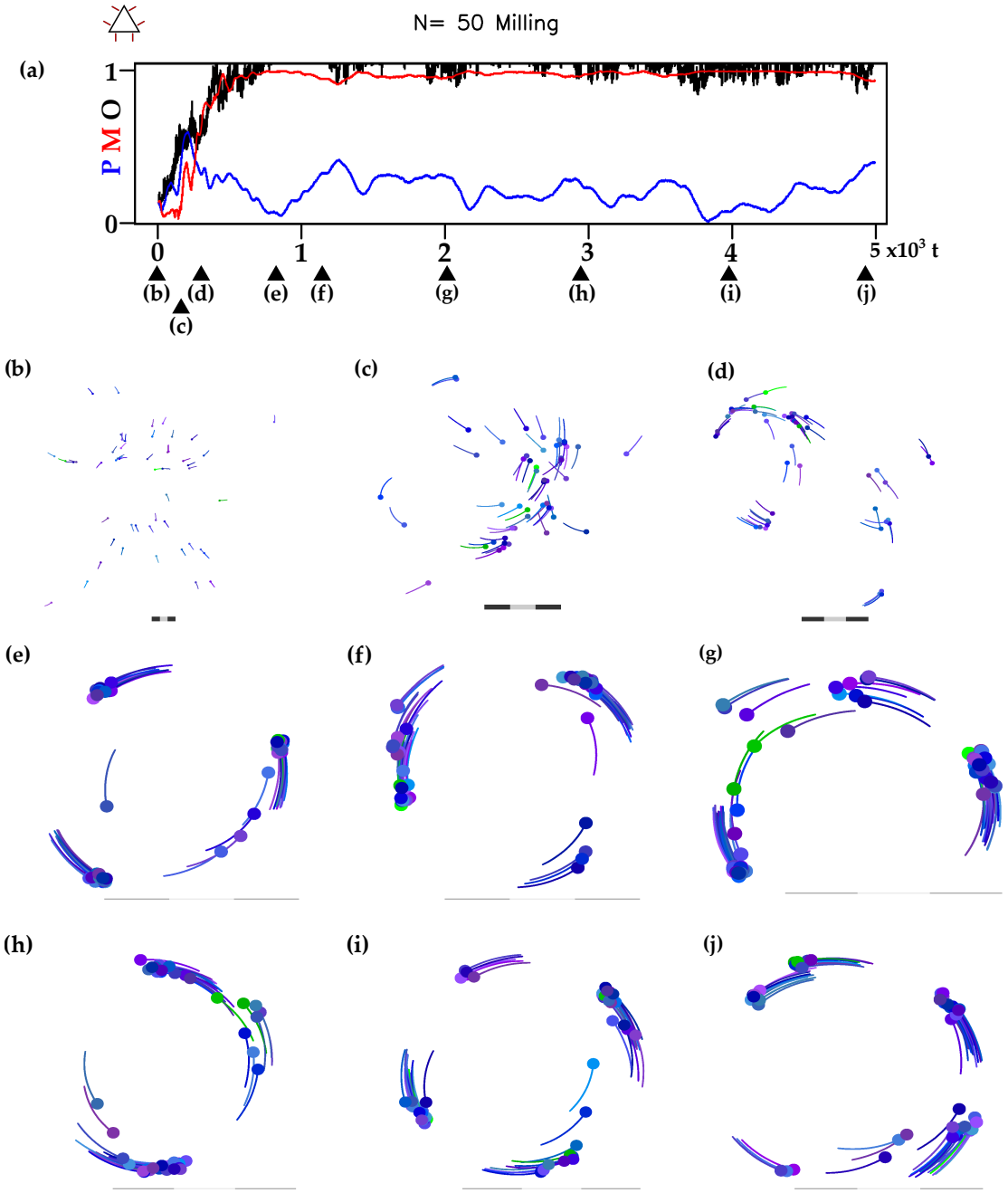

Figure A.25.: Milling phase reproduction for a flock size of 50 individuals for the Omniscient model. k_η =0.06 k_η =0.06 k_η =0.01 and maximum blind angle .

B. Supplementary information: Visual collective behaviors on spherical robots

B.1. Image Processing Technical information

The camera details are presented on Table SB.1 The detailed dual image segmentation

Resolution	1920x1200
Frame rate	168fps
Pixel Size	5.86uM
Sensor Type	Sony IMX174
Optical Size	1/1.2"
Shutter Type	Global Shutter
Shutter time	52us 1s
ADC Bit Depth	10bit
Pixel Bit Depth	8bit, 10bit
Digital gain	0dB 24dB
Pixel Data Format	Bayer RG8 / Bayer RG10
Lens mount	type-C
Power Consumption	<2.7W@5V

Table B.1.: Camera: *Daheng Imaging MER2-230-168U3C* specifications.

example is presented on Fig. SB.1.

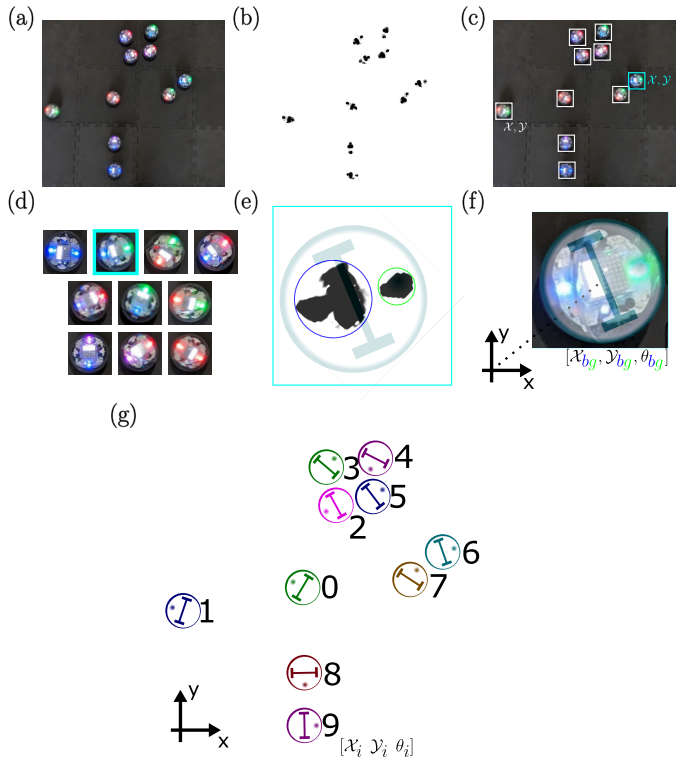

Figure B.1.: Image Processing. a) Representation of Raw image, b) Binary segmentation based on luminosity c) Coordinates extraction for all robots on the experiment, d) Image segments of each of the isolated robots, e) Segmentation and color identification of the front LED and LED matrix + back LED, e) robot identification (XY-coordinates + heading) based on color codification, g) Virtual representation of the identified robots, using with their own coordinates and heading..

B.1.1. Feature Matching Optic Flow

Fig. B.2 showcases examples of what real world each of the components of the optic flow would look like. It also correlated each with a binary panorama equivalency. Finally it shows a mixed case that would be a normal optic flow generation for both image and 1D cases.

For the feature matching optic flow the gradient generated when the amount of retinal object differ between consecutive time steps is shown on Fig. B.3.

B.2. Digital implementation of the supplemented visual model

The model presented on Eqs. (3.2)-(3.7), are the integral form of the model that use of the visual panorama. $\mathcal{V}_i(\phi, t-\tau)$ is a piece-wise constant function that can be

expressed as a set unit boxcar functions (one for each retinal object).

Each retinal object can be described as a rising $\varphi_{\Uparrow k}$ and its corresponding falling edge $\varphi_{\Downarrow k}$. Furthermore, from these edges, we can compute the shade's mid-point $\varphi_k = (\varphi_{\Uparrow k} + \varphi_{\Downarrow k})/2$, and the retinal's object half-width $\Delta\varphi_k = (\varphi_{\Downarrow k} - \varphi_{\Uparrow k})/2$. The apparent distance, $\mathscr{R}_i(\phi, t-\tau)$ is defined as $\mathscr{V}_i(\phi, t-\tau)$ weighted by $\frac{a}{\sin\Delta\varphi_k}$.

This means that $\mathscr{R}_i(\phi, t-\tau)$ for a binary $\mathscr{V}_i(\phi, t-\tau)$, is a function independent of ϕ and is only be dependent on the k-retinal's object half-width ($\mathscr{R}_k(\Delta\varphi_k)$). Similarly, the avoidance function can be expressed as a function of $\Delta\varphi_k$, and is independent of ϕ. We add the subscript k, which identifies a single retinal object on the binary panorama of the individual i. Note that a retinal object k is different from an individual j, as in a retinal object there can be occlusion or aggregation of any j-particles on the i-th binary panorama. Then, we can rewrite Eq. (3.4) as:

$$\omega_\odot(t) = \left\langle \sum_k \mathscr{R}_i^2(\Delta\varphi_k, t-\tau)\mathscr{A}(\Delta\varphi_k, t-\tau) \int_{\varphi_{\Uparrow k}}^{\varphi_{\Downarrow k}} b_\epsilon(\phi, t-\tau)\sin\big(\phi, t-\tau\big)\, d\phi. \right\rangle \tag{B.1}$$

where $b_\epsilon(\phi, t-\tau) = 1 + \epsilon\cos\big(\phi, t-\tau\big)$. And developing the integral we get

$$\omega_\odot(t) = \langle \sum_k 2\mathscr{R}_i^2(\Delta\varphi_k, t-\tau)\mathscr{A}(\Delta\varphi_k, t-\tau)\sin\big(\Delta\varphi_k, t-\tau\big)\sin\big(\varphi_k, t-\tau\big) \big(1 + \epsilon\cos\big(\varphi_k, t-\tau\big)\cos\big(\Delta\varphi_k, t-\tau\big)\big)\rangle \tag{B.2}$$

And here $\mathscr{R}_i(\Delta\varphi_k, t-\tau) = \frac{a}{\sin(\Delta\varphi_k, t-\tau)}$, Resulting in:

$$\omega_\odot(t) = \left\langle \sum_k \frac{2a^2\left(1 + \epsilon\cos\big(\varphi_k, t-\tau\big)\cos\big(\Delta\varphi_k, t-\tau\big)\right)}{\sin\big(\Delta\varphi_k, t-\tau\big)} \sin\big(\varphi_k, t-\tau\big) \right\rangle \tag{B.3}$$

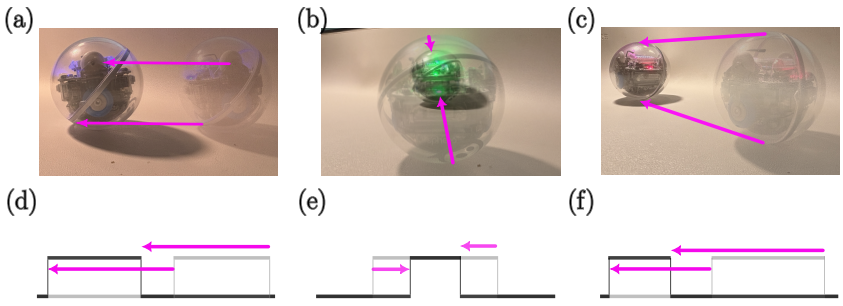

Figure B.2.: Optic flow. a) Purely translational, b) Purely Convergent, c) Mixed Case, d) 1-D feature matching purely translational, e) 1-D feature matching purely convergent, f) 1-D feature matching mixed case.

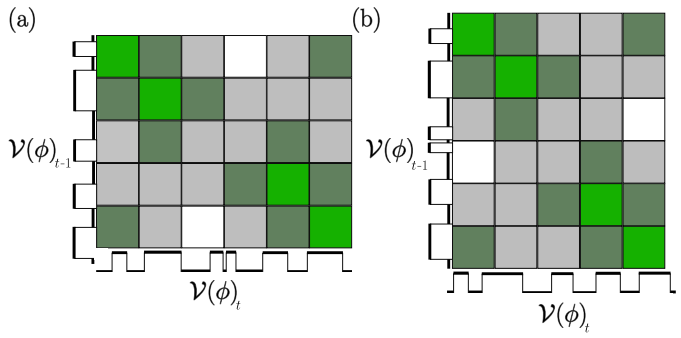

Figure B.3.: Retinal object matching between consecutive time steps. a) Separation of retinal object, b) Merging of retinal objects.

By substituting the $\langle\cdot\rangle$ normalization the implemented equation is:

$$\omega_{\odot}(t) = \frac{\sum_k \frac{2a^2(1+\epsilon\cos(\varphi_k,t-\tau)\cos(\Delta\varphi_k,t-\tau))}{\sin(\Delta\varphi_k,t-\tau)}\sin\big(\varphi_k,t-\tau\big)}{\sum_k \frac{2a^2(1+\epsilon\cos(\varphi_k,t-\tau)\cos(\Delta\varphi_k,t-\tau))}{\sin(\Delta\varphi_k,t-\tau)}} \tag{B.4}$$

The next step is to do the same for the alignment component. First, we replace Eq. (3.7) in Eq. (3.5) resulting in:

$$\omega_{\parallel}(t) = \int_{-\pi}^{\pi} U^{-1}(\overline{\mathscr{O}}_i(\phi,t-\tau)\cos\big(\phi,t-\tau\big) - \mathscr{D}_i(\phi,t-\tau)\sin\big(\phi,t-\tau\big))b_{\epsilon}(\phi,t-\tau)d\phi \tag{B.5}$$

Note, that the visual alignment can be also be defined as a vector with an azimutal and radial component, which is defined as:

$$\omega_{\parallel}(t) = \langle\omega_{\parallel,\phi}(t) + \omega_{\parallel,r}(t)\rangle \tag{B.6}$$

$\mathscr{D}_i(\phi,t-\tau)$ and $\overline{\mathscr{O}}_i(\phi,t-\tau)$, similarly to $\mathscr{R}_i(\phi,t-\tau)$, are piece-wise constant function which are independent of ϕ but dependent on $\Delta\varphi_k$.

Taking Eq. B.5 and Eq. B.6:

$$\omega_{\parallel,\phi}(t) = \sum_k \mathscr{O}_i(\Delta\varphi_k,\phi,t-\tau)\left(\int_{\varphi_{\Uparrow k}}^{\varphi_{\Downarrow k}} \cos\big(\phi,t-\tau\big)b_{\epsilon}(\phi,t-\tau)d\phi\right) \tag{B.7}$$

$$\omega_{\parallel,r}(t) = \sum_k -\mathscr{D}_i(\Delta\varphi_k,\phi,t-\tau)\left(\int_{\varphi_{\Uparrow k}}^{\varphi_{\Downarrow k}} \sin\big(\phi,t-\tau\big)b_{\epsilon}(\phi,t-\tau)d\phi\right) \tag{B.8}$$

Eq. B.8 is the exact form as Eq. B.1 with the weight of the integral being $\mathscr{D}_i(\Delta\varphi_k,t-\tau)$ instead of $\mathscr{R}_i^2(\Delta\varphi_k,t-\tau)\mathscr{A}(\Delta\varphi_k,t-\tau)$ resulting in:

$$\omega_{\parallel,r}(t) = \sum_k \frac{-2\mathscr{D}_i(\Delta\varphi_k,t-\tau)}{\mathscr{R}_i(\Delta\varphi_k,t-\tau)}\sin\big(\varphi_k,t-\tau\big)\big(1+\epsilon\cos\big(\varphi_k,t-\tau\big)\cos\big(\Delta\varphi_k,t-\tau\big)\big) \tag{B.9}$$

As for Eq. B.7 the developing the integral and assuming $\Delta\phi \approx \sin\Delta\phi$ we have:

$$\omega_{\parallel,\phi}(t) = \sum_k \frac{\mathscr{O}_i(\Delta\varphi_k, t-\tau)}{\mathscr{R}_i(\Delta\varphi_k, t-\tau)}(2\cos(\varphi_k, t-\tau)(1+\epsilon\cos(\Delta\varphi_k, t-\tau)\cos(\varphi_k, t-\tau)) + \epsilon(1-\cos(\Delta\varphi_k, t-\tau))) \quad \text{(B.10)}$$

Assumming that $\cos\Delta\varphi_k \approx 1$ and using Eq. 1f, then Eqs. B.9-B.10 become:

$$\omega_{\parallel}(t) \approx \left\langle \sum_k 2\frac{\mathbf{e}_i(t-\tau)\times\mathbf{V}_{ik}(t-\tau)}{U\mathscr{R}_i^2(\Delta\varphi_k, t-\tau)} b_\epsilon(\varphi_k, t-\tau) \right\rangle \quad \text{(B.11)}$$

Susbtituting the normalization and the blindspot, Eq. B.11 becomes:

$$\omega_{\parallel}(t) \approx \frac{\sum_k 2\frac{\mathbf{e}_i(t-\tau)\times\mathbf{V}_{ik}(t-\tau)}{U\mathscr{R}_i^2(\Delta\varphi_k, t-\tau)}(1+\epsilon\cos(\varphi_k, t-\tau))}{\sum_k 2\frac{\|\mathbf{V}_{ik}(t-\tau)\|}{U\mathscr{R}_i^2(\Delta\varphi_k, t-\tau)}(1+\epsilon\cos(\varphi_k, t-\tau)} \quad \text{(B.12)}$$

B.3. Observed phases

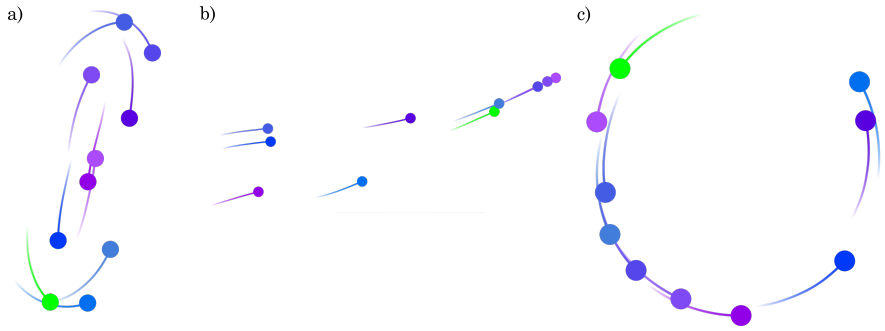

Figure B.4.: Illustration of the phases observed for the raw original model Castro et al. 2024b. We observe three phases when varying the other two parameters: (a) Swarming ($k_\odot = 0.15$, $k_\parallel = 0$) (see Svideo 13); (b) Schooling ($k_\odot = 0.07$, $k_\parallel = 0.3$) (see Svideo 14); (c) Milling ($k_\odot = 0.07$, $k_\parallel = 0.05$) (see Svideo 15).

B.4. Occlusion Analysis

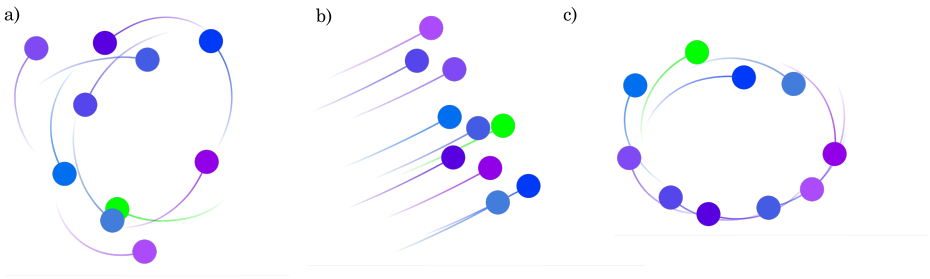

Figure B.5.: Illustration of the phases observed for the visual model supplemented only with avoidance. We observe three phases when varying the other two parameters: (a) Swarming ($k_\odot = 0.07$, $k_\parallel = 0$) (see Svideo 16); (b) Schooling ($k_\odot = 0.15$, $k_\parallel = 0.3$) (see Svideo 17); (c) Milling ($k_\odot = 0.07$, $k_\parallel = 0.05$) (see Svideo 18).

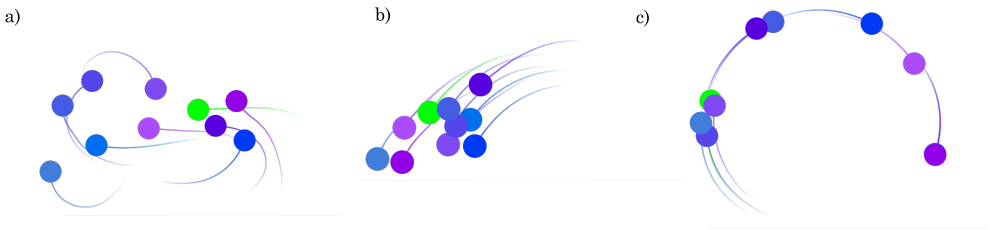

Figure B.6.: Illustration of the phases observed for the visual model supplemented only with the avoidance and the visual anchor. We observe three phases when varying the other two parameters: (a) Swarming ($k_\odot = 0.3$, $k_\parallel = 0$) (see Svideo 19); (b) Schooling ($k_\odot = 0$, $k_\parallel = 0.3$) (see Svideo 20); (c) Milling ($k_\odot = 0.07$, $k_\parallel = 0.3$) (see Svideo 21).

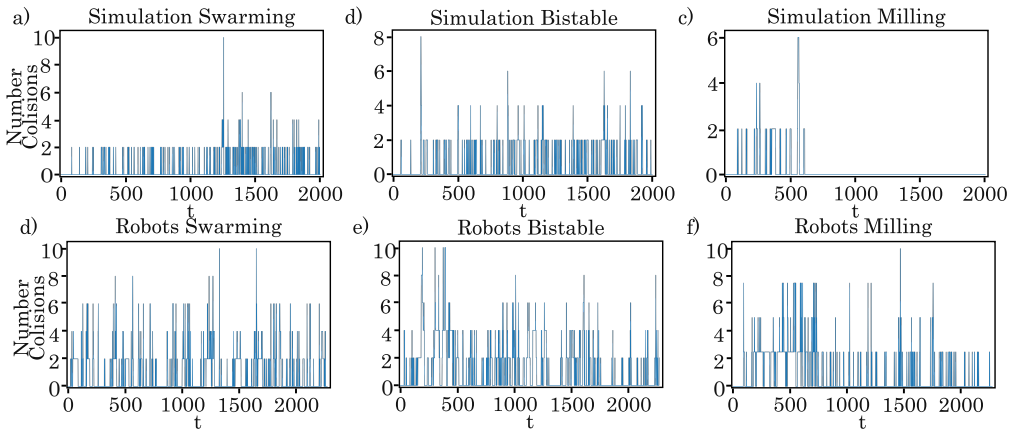

Figure B.7.: Number of collisions
Number of collisions during the evolution with a set of parameters (a) Swarming simulation; (b) Schooling simulation; (c) Milling simulation; (d) Swarming robot-in-the-loop; (e) Schooling robot-in-the-loop; (f) Milling robot-in-the-loop.

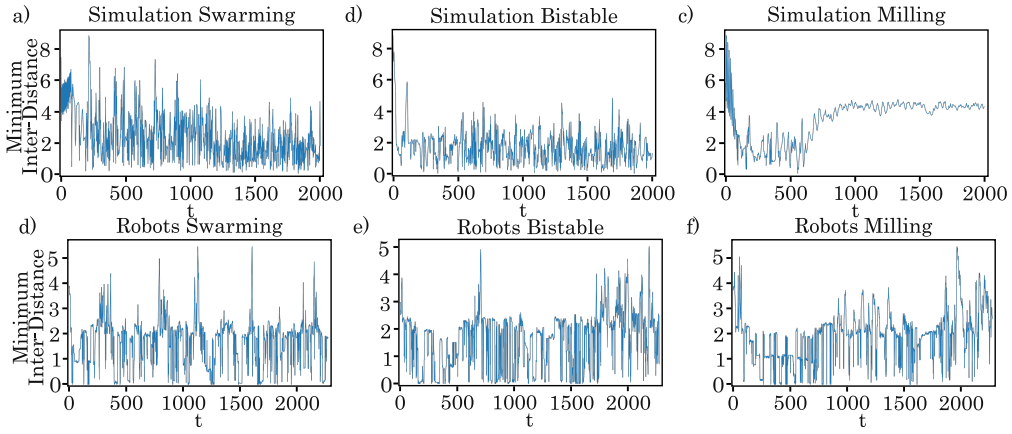

Figure B.8.: Minimum inter-distance
Minimum inter-distance found during the evolution with a set of parameters (a) Swarming simulation; (b) Schooling simulation; (c) Milling simulation; (d) Swarming robot-in-the-loop; (e) Schooling robot-in-the-loop; (f) Milling robot-in-the-loop.