UNIVERSITAT POLITÉCNICA DE CATALUNYA

DOCTORAL THESIS

Collective Motion and Collective Decision-Making in Animal Groups: From Schooling Fish to Swarming Robots

---

Author:
Julia MÚGICA GALLART

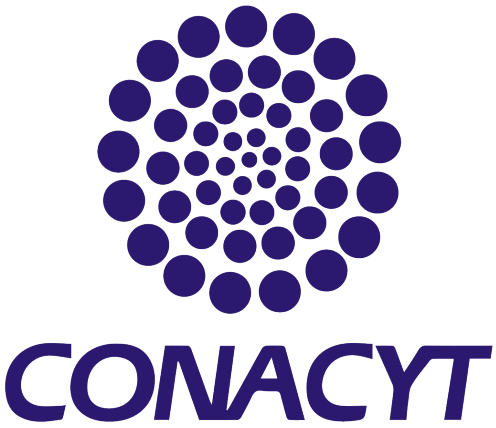
Supervisors:
Dr. Romualdo PASTOR-SATORRAS
Dra. M. Carmen MIGUEL LOPEZ

A thesis submitted in fulfillment of the requirements for the degree of Doctor of Philosophy

in the

Department of Physics

March 16, 2023



Scholarship number
625639

Abstract

Collective behavior in animals is ubiquitous in nature. It emerges from the self-organization of interacting individuals within a group. What makes collective behavior so interesting is that the behavior of the group as a whole cannot be predicted from the actions of its individual members alone. In recent decades, there has been a growing interest in understanding how collective behavior emerges from individual interactions. This research has revealed that self-organization in biological systems share common functional properties that can trigger behavioral transitions that endow collectives with adaptive capabilities.

The main objective of this work is to explore the underlying mechanisms from which complex collective behavior can emerge and adapt. We concentrate our analysis in the collective motion of schooling fish and consensus reaching in swarming robots in a model of honeybees nest site selection.

We study the collective motion of black neon tetra (*Hyphessobrycon herbertaxelrodi*) by recording their movements in an experimental tank and using a custom tracking software to extract individual trajectories. Our analysis of the data reveals a strong correlation between swimming speed and collective ordering. We explore this correlation in the Vicsek model, a self-propelled particles model, and discover that the system undergoes a disorder-to-order phase transition with an increase in particles' speed. By modifying the Vicsek model to include periodically oscillating speed, we are able to replicate the black neon tetra's type of locomotion. This suggests that fish can reduce the effect of noise by increasing their speed and favoring alignment forces. Modulating directional synchronization through speed has adaptive benefits, enabling faster and more efficient transfer of information.

We also investigate the presence and characteristics of avalanches in our fish school, which occur when fish abruptly rearrange their direction of motion. We observe that the size and duration of these avalanches follow a scale-free pattern similar to self-organized critical systems. Additionally, certain fish are more likely to initiate avalanches, acting as effective leaders. By incorporating a global leader that periodically changes direction, we replicate in the Vicsek model the intermittent

scale-free behavior observed in fish schools. Our findings highlight the importance of individual variability in triggering behavioral cascades.

To analyze collective decision-making we study a decentralized model that simulates honeybees' nest site selection process. The model integrates independent site discovery and communication to reach consensus on the best site. We thoroughly characterize the model's parameters in a mean-field approximation and test its feasibility by adding space, locality, and situatedness to the dynamics by using kilobots, small swarming robots capable of exchanging information with neighboring robots. Remarkably, the level of stationary consensu s in our results from kilobots closely matches the consensus stationary values predicted by the mean-field approximation. To explain this result, we examine the connectivity of the kilobots system in our experimental arena, and demonstrate that, from local interactions and mobility, kilobots form a percolating communication network from which they transmit information as efficiently as in the mean-field approximation. The kilobots' movement effectively induces an extension of the interaction radius in respect to the static scenario, which is especially important in the high communication regime. Our results demonstrate that this model can explain honeybees consensus reaching on the best available site by local dynamical interactions.

Our findings inspire further exploration in individuals variability in self-organizing systems, with the aim of developing more comprehensive models and expanding knowledge of the universality of collective behavior.

Resum

El comportament col·lectiu dels animals és ubic a la natura. Sorgeix de la interacció entre individus d'un grup. Tot i així, no pot ser predit només a partir de les accions individuals dels seus integrants. En les darreres dècades, hi hagut un interès creixent en entendre com el comportament col·lectiu emergeix de les interaccions individuals. Aquesta recerca ha revelat que l'autoorganització en sistems biològics comparteix propietats funcionals comunes que poden desencadenar transicions que doten als col·lectius de capacitats adaptatives.

Aquest treball es centra en l'estudi dels mecanismes pels quals els comportaments col·lectius complexes sorgeixen i s'adapten. Ens centrarem en l'anàlisi del moviment col·lectiu de bancs de peixos i el consens en eixams de robots que modelen el procés amb el qual les abelles escullen l'ubicació d'un nou rusc.

Estudiem el moviment col·lectiu del peix tetra negre neon (*Hyphessobrycon herbertaxelrodi*) amb un programari personalitzat que extreu les trajectòries individuals dels individus d'un banc a partir d'enregistraments d'un aquari experimental. Els resultats mostren una forta relació entre la velocitat a la que neden i l'ordre col·lectiu. Explorem aquesta correlació en el model de Vicsek, que demostra la transició de fase, des del desordre fins a l'ordre, amb l'augment de la velocitat de les partícules. Modificant el model de Vicsek per incloure una oscil·lació de velocitat periòdica, podem replicar la locomoció del tetra negre neon. Això suggereix que els piexos poden reduir l'efecte del soroll augmentant la velocitat i afavorint les forces d'alineament. Modulant la sincronització direccional amb la velocitat, els peixos poden intercanviar informació de manera més ràpida i eficient, la qual cosa té beneficis adaptatius.

També investiguem la presència i característiques de les allaus als bancs de peixos quan canvien sobtadament de direcció. S'observa que la mida i durada d'aquestes allaus segueix un patró lliure d'escala, similar al de sistemes crítics autoorganitzats. Adicionalemnt, alguns peixos són més propensos a desencadenar allaus, actuant efectivament com a líders. A l'incorporar un líder global que periòdicament canvia de direcció, repliquem – en el model de Viscek – el comportament lliure d'escala obervat en els bancs de peixos.

Per analitzar le presa de decisions col·lectives, estudiem un model descentralitzat que simula el procés pel qual les abelles escullen l'indret per un nou rusc. El model integra la cerca independent d'ubicacions i la comunicació per arribar a un consens pel millor lloc. Carecteritzem els paràmetres del model amb una aproximació de camp mitjà i provem la seva viabilitat afegint espai, localitat i contextualització a la dinàmica utilitzant kilobots, robots d'eixam capaços d'intercanviar informació amb robots veïns. El nivell de consens estacionari als nostres resultats amb kilobots s'assembla notablement als valors del consens estacionari obtinguts per aproximació de camp mitjà. Per explicar aquest resultat, examinem la connectivitat del sistema de kilobots a la nostra arena experimental, i demostrem que – a partir de les interaccions locals i la mobilitat – els kilobots formen una xarxa percolada de comunicació des de la que transmeten informació tan eficientment com en l'aproximació de camp mitjà. El moviment dels kilobots indueix efectivament una extensió del radi d'interacció respecte a l'escenari estàtic, especialment important en el règim d'alta comunicació. Els nostres resultats demostren que aquest model pot explicar com les abelles arriben a un consens sobre la millor situació per al nou rusc amb interaccions dinàmiques locals.

Els nostres descobriments inspiren una exploració més profunda en la variabilitat dels individus en l'autoorgantització de sistemes, amb l'objectiu de desenvolupar models més extensos i expandir el coneixement sobre l'universalitat del comportament col·lectiu.

Acknowledgements

First of all, I would like to express my gratitude to Carmen and Romualdo, my supervisors, for welcoming me into their research group and providing me with the opportunity to explore the world of physics, coming from a biology background. Thank you for the enlightening discussions we had, and for giving me the opportunity and insight to your thought processes and perspectives on scientific challenges. I have learned a great deal from both of you.

I would also like to acknowledge the scientific committee of this thesis. I know that it is quite lengthy, and I sincerely appreciate the time, patience, and interest you have dedicated to reading and engaging with this work. Thank you.

I am also thankful for David and Ezequiel and their contributions to the discussions on kilobots. Your involvement was crucial for the work to arrive to a successful outcome. Specifically to David; this has been an journey and I can only be grateful that you were the person that I was able to share it with. In the context of this project, I would be remiss in not mentioning Lina, Andrés and Iván for their invaluable contribution in developing an experimental setup for the robots. Your involvement was not only instrumental but also made the process enjoyable.

I would also like to extend my sincere thanks to Jordi, Andreu, and Jaume. Your camaraderie made me feel like I had a team to share my PhD experiences with. Although it came later in my PhD journey, having a student group made all the difference in the world. Jordi and Andreu, our discussions of fish were always very insightful. I have deep admiration for the way both of you continuously question results and go beyond what is expected. Within this project, thanks should also go to the *Love your data* team, which made the analysis of our fish possible with their contributions to the tracking software.

I would also like to thank Eli, Vicenc, and Francesc for introducing me to this research topic while doing my Masters studies, and giving me the opportunity to work with black neon tetras and zebrafish in their laboratory. Through this experience, I gained valuable insights into working with animals for research purposes, and Eli's compassionate approach towards them reminded me that more than anything, living beings are to be treated with care and respect.

I would like to express my profound gratitude to Rebekah for her contribution to the writing of this work. Her thorough reading of the entire thesis and review of the English was instrumental in clarifying convoluted ideas.

On a personal level, I wish to acknowledge the significance of my incredible support network throughout this entire process.

First of all, this achievement would not have been possible without the unwavering backing of my family. Every member of my family has taught me something valuable that I have been able to apply in my life and studies over the years.

Moreover, I am deeply grateful for my beautiful friends. The ones on the other side of the world and the ones living a few blocks away. I cannot overstate how fortunate I feel to know you all. You have listened to me exhaustively, shared in my joys and my pains, and allowed me to reciprocate. I thank you for journeying through life together and always having my back. On the same note, a deep thank you to Laura and the *Re-crearnos* group. Through our work together I have learned to listen to myself and value my own dignity. Also I would like to thank the *Axolot* collective. You have showed me how people from such different backgrounds and skills can create something truly remarkable.

Lastly, this journey would not have been possible without Iván (and Manolla). Sharing a home with you has been an incredible experience that I am truly grateful for. You inspire me every day, and your unwavering support has been a source of comfort and joy through both good times and bad. I love you, and thank you.

Contents

List of Figures

List of Tables

For Manolla

Preface

I can pin point the exact moment when I became captivated by the concept of emergence in collective behavior. Back in 2013 I was studying Biology at the National Autonomous University of Mexico (UNAM). I was doing my thesis on the genetic flux between multiple tree populations in the Transvolcanic Mexican Belt and while it was certainly an interesting topic, I already knew I was not going to pursue a research careeer on evolutionary genetics. During a chat with a dear professor of mine, Pablo Padilla, he shared with me that he was working on a research project about ant foraging behavior. That's when he introduced me to a Netlogo simulation that demonstrated how positive feedback in ant pheromone trails could help ants locate the shortest path to food without any central direction.

I was completely mesmerized by the fact that without any planning such a complex behavior could emerge and, moreover, modeled with relatively simple instructions. But what truly captured my imagination was the realization that this type of emergent behavior was not unique to ants, but a common phenomenon in nature. This led me to the understanding that the emergence of collective behavior might be more of a universal property of living beings, and hence, not dependent on any specific cognitive abilities of a particular species. Collective behavior could potentially rely on the simple interactions between individuals. As individuals within a group interact with one another, their behavior shifts in unpredictable ways, leading to emergent complex dynamics, making collective behavior an endlessly fascinating subject to study.

Of course, the months after that conversation I started consuming all types of Complexity Science content. I enrolled myself in many of the online courses offered by the Santa Fe Institute and learned concepts such as non-linear dynamics, scaling behavior, self-organization, and so on. I also started playing with Netlogo and, with that, I realized I loved programming.

When I moved to Barcelona, I got into the "Research in Behavior and Cognition" master studies in the Faculty of Psychology of the University of Barcelona (UB). That experience introduced me to one of the many perspectives for studying complex phenomena. In psychology, the primary goal of studying collective behavior is to comprehend the social interactions that give rise to it and the particular group behavior that emerges from such interactions. During my studies, I had the opportunity to explore interesting examples of primates social organization through simple interactions. I was also introduced to black neon tetras and zebrafish. These small freshwater fish were the focus of the research group where I completed my master's final work, and they provided a unique opportunity to investigate collective motion differences between them.

After finishing the master's degree, I wanted to deepen my understanding of the principles of animal collective behavior and its possible universality. To achieve this, I pursued a PhD in Applied Physics at the Universitat Politècnica de Catalunya (UPC), where I aimed to combine empirical work of animal emergent collective behaviors and theoretical approximations that could shed light into the underlying interactions between animals and their properties. Statistical physics is a powerfull framework for studying collective behavior as it deals with systems of many interacting units to understand their macroscopic behavior. The tools and perspective this field offers has largely enriched our understanding of complex behavior in the last decades.

The work in the PhD was quite challenging for multiple reasons. As someone who came from a different field, I had to deal with the discomfort of learning a new scientific "language". Interdisciplinary research involves effective communication to tackle the different assumptions and ways of thinking about a problem, and that in itself takes time to build. In addition, interdisciplinary research requires a significant investment of time and resources. For example, we had to design and set up many of the experimental requirements to empirically work with collective systems. These efforts can take years and can end up occupying a quarter of a slide in your presentation and placed in the last page of the scientific paper, only thoroughly read by the extremely curious.

Despite the challenges, interdiscipline has the amazing quality of promoting innovative solutions to a problem, and we got to practice this more than once. It broadens your perspective by exposing you to different ways of thinking. This is a quality that I place very high on the top of my priorities in both personal and professional development. Moreover, addressing complex problems from an interdisciplinary perspective has the potential to develop a more comprehensive understanding of the phenomena and offer effective solutions to complex issues. As in complex collective behavior, individual variability sometimes is the spark that ignites a whole different research outcome. On that note, I can only remind the reader that this work is a collective effort that could only have emerged from the involvement and interaction of multiple people.

During the years we worked on our project, there was a growing interest in animal collective behavior among researchers. This was in part due to the development of new technological tools that enabled the collection of large amounts of behavioral data, allowing for more detailed analyses of animal behavior. Our project is part of this larger effort to understand collective behavior and its underlying mechanisms. Moreover, in today's digital age, technological advancements are accelerating changes to our social systems, and we are struggling to predict and respond to these changes. Therefore, understanding how collective behavior emerges in different contexts is more important than ever.

1 Introduction

1.1 Animal Collective Behavior

Collective behavior is a complex and a widespread phenomenon in biological systems emerging on multiple scales, hierarchies and complexity levels (Vicsek and Zafeiris, 2012). It ranges from bacterial colonies and migrating cells (Deisboeck and Couzin, 2009) to the spread of rumors within human beings (Miller, 2013). Within animal species, fascinating collective behavior is displayed with the trails of foraging ants (Wilson, 1971), the movement of flocks of birds (Balda and Bateman, 1971) and light syncronization in sparkles of fireflies (Sarfati et al., 2021; see Figure 1.1).

Interestingly, within these examples of collective behavior it has been observed that there is neither a leader nor external cues guiding the group towards a collective pattern. Collective behaviors emerge from **self-organization** by individuals interacting locally and exchanging information, and not as a product of a central control (Sumpter, 2006). Even in animals with strong social hierarchies, such as societies of primates or honeybee colonies, the behavior of the collective is not dictated by a single leader (Wild et al., 2021). Instead, it has been shown that the collective behaviors exhibited in bees and other social animals are guided by consensus mechanisms (Romanczuk and Daniels, 2023).

Through local interactions, each individual action is influenced by others within the group, instigating it to behave differently as it would solitarily. In other words, the global behavior cannot be directly inferred from the knowledge of the individual components (Vicsek and Zafeiris, 2012) as *the whole is more than the sum of its parts*. A direct consequence of this property in self-organized systems is its deviation from the *central limit theorem*. This mathematical theorem states that if a system consists of many individuals acting independently, and each individual contributes with a

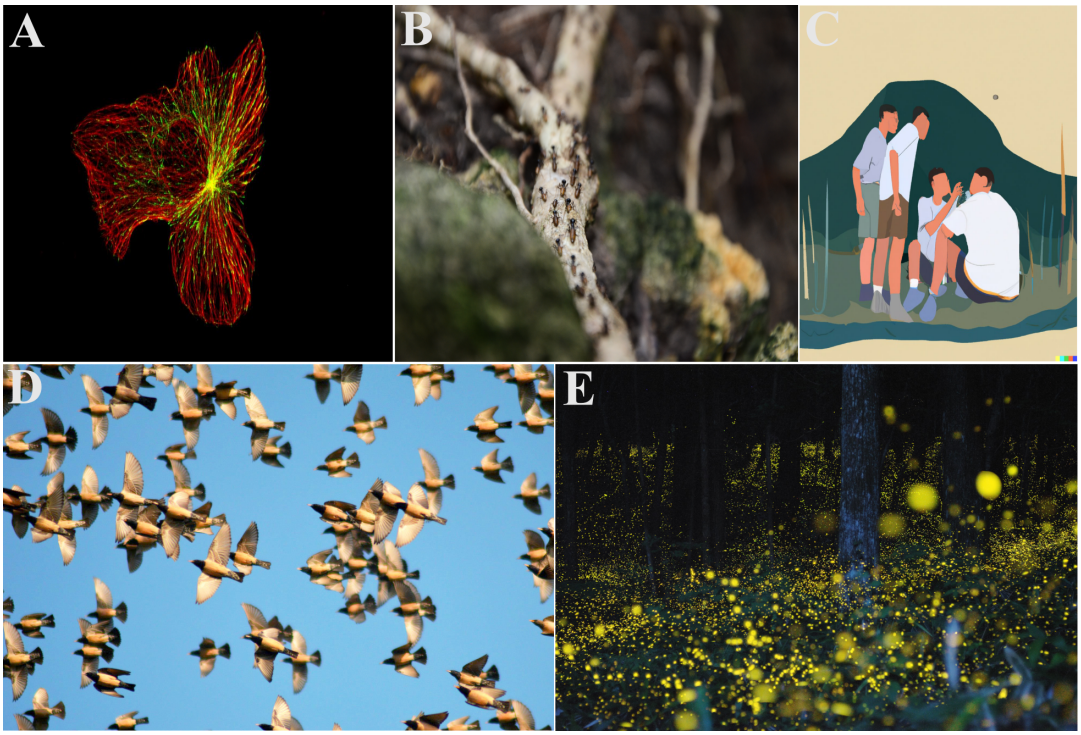


Figure 1.1: Collective behavior in different biological systems. **A: Migrating cancel cell** (by Brian6122, licence CC BY 2.0)**B: Ant trails** (by Theo Crazzolara under CC BY 2.0) **C: Rumours** (Image created by DALL-E AI interface with the input text "minimalist painting of group of people together sharing secrets and rumours in nature") **D: Flock of Rosy Starlings *Pastor roseus***, near Delhi (by Thomas Avey, licence CC BY-SA 4.0) **E: Japanese fireflies** (by Tsuneaki Hiramatsu, licence CC BY-NC-ND 2.0)

randomly distributed quantity to the total output, then the resulting sum tends toward a normal distribution (Sornette, 2006). This is not evident in systems where collective patterns emerge from individuals' interactions, so they can be explained in terms of the deviation they exhibit from a normal distribution.

Since information is transmitted locally through the group and there is no centralized control, individuals within the collective may have varying informational status and unaware of intelligence gathered by others in the collective, such as a food resource or a threat (Couzin, 2007). Nonetheless, individuals living in a group must often make rapid decisions regarding movement or which type of behavior to adopt in order to face the challenges imposed by their external context. All biological systems require the capacity of rapid adaptive change. **Transitions** between different types of behavior are commonly observed in nature. They occur when animals shift from one collective behavior to another spontaneously or in response to a stimulus, such as groups of fish that change their cohesion and synchronization when they find themselves under a shade (Ribeiro et al., 2022). In the last decades

experimental and theoretical studies have begun to elucidate how collective behavior emerges from the repeated individual interactions and what are the mechanisms that trigger behavioral transitions (Tunstrøm et al., 2013; Calovi et al., 2014). Interestingly, they have revealed that self-organization shares similar functional properties in a wide range of animal groups, including humans.

From a biological point of view, the study of collective animal behavior allows us to relate two different approaches to these analysis of natural phenomena. A mechanistic perspective seeks explanations about how animals interact to produce global patterns. It focuses on identifying the underlying communication mechanisms from which a collective pattern emerges. Alternatively, a functional perspective concentrates on understanding why evolution favored a collective behavior through natural selection. Although in other areas of biology these two approaches can be studied separately, in collective animal behavior they are interdependent. We cannot decipher why it is that collectiveness evolved without understanding the mechanisms that generate it, and we cannot examine mechanisms without limiting potential explanations to a viable adaptive scenario (Sumpter et al., 2008).

Collective animal behavior is a central phenomenon in behavioral biology studies, but it is also relevant in other scientific areas where collective behavior and self-organization have been observed, including physics (Boolchand et al., 2005), economics (Cont and Bouchaud, 2000), social sciences (Helbing, 2012), artificial intelligence (Thrun and Ultsch, 2021) and robotics (Pfeifer et al., 2007). From this perspective, animal groups are emblematic examples of collective behavior and have triggered a growing multi-disciplinary interest and application. Particularly, collective behavior is a central concept in statistical physics, where the emergence of order and the transitions from one global state to another in collective systems have been intensively investigated in the last fifty years. Because of this, physics is regarded as a source of inspiration, and its tools and methodologies have been recently applied to problems of collective phenomena in other fields (Giardina, 2008).

Mainly, collective systems in physics are relevant for other areas because they consist of individual components, such as particles or magnetic moments, that by means of local interactions can generate an *ordered* state with collective global properties. Moreover, the underlying fundamental principles of collective phenomena

are usually generalized, and therefore nearly independent of the specific details relative to the behavior of individual components, so they are technically transferable between systems of different nature (Bode et al., 2011). This characteristic is known as universality (Camazine et al., 2001a) and its possible existence in animal collective mechanisms have generated a growing interest in animal collective behavior by both statistical physicists and theoretical biologists, as well as by experimental scientists (Deutsch et al., 2012).

From a theoretical point of view, the methodological tools of analysis borrowed from physics include observables to quantify the internal structure of a collective system, specially their dynamical features and susceptibility to perturbations (Giardina, 2008), and the generation of models of self-organized behavior, that have proven useful for extracting connections between the interactions of individuals and the collective patterns that emerge from them, that would otherwise be difficult to detect (Giardina, 2008).

Models of collective animal behavior are constructed assuming some behavioral rules at the individual level from which the global behavior can be analyzed. By varying the parameters involved in local interactions the different resulting behaviors at the global level can be explored (Vicsek et al., 1995; Czirók et al., 1997; Parrish et al., 2002; Couzin, Krause, et al., 2003). Depending on the goal of the researcher, the individual rules of interaction can be modified for sophistication. Often biologists want to describe specific animal groups while physicists address more conceptual questions and seek to build the *minimal* set of rules needed to produce collective behavior (Giardina, 2008). Normally self-organized models cannot be solved exactly by coupled dynamical equations. Rather, these models are easily implemented as numerical simulations. Nonetheless, there are instances where the collective behavior to be explained can be addressed by a simpler mathematical formulation that provide quantitative predictions in terms of a small number of parameters (Giardina, 2008).

From an empirical point of view, last decades have witnessed an explosion in experimental activity (Beckers et al., 1992; Beekman et al., 2001; Dussutour et al., 2008). Advances in image and tracking technology have enabled experimental scientists to obtain high precision data and make an accurate analysis of individuals interactions

and characterize grouping behavior. From these data it is possible to compare the observations with theoretical models, and to infer the underlying mechanisms of communications in collective behavior events (Vicsek and Zafeiris, 2012).

Observing the distinctive ways in which individuals interact is pivotal in representing the source of the diversity of structures and dynamics of collective motion in nature. In both theoretical and experimental works it has been shown that animal groups can be subject to both behavioral group and individual variations (Couzin et al., 2002; Nagy et al., 2013; Tunstrøm et al., 2013; Herbert-Read et al., 2015; Biro et al., 2016). Group variation refers to the group's capacity to exhibit multiple stable collective states without necessarily changing the rules of interaction, but by changing the frequency and the force by which individuals interact. The latter refers to individuals' probability of responding to the movements of their neighbors. Group variations can occur as a response to an external perturbation, to internal states, or to different stages of development. Individual variation expresses in the difference in the rules of interaction within a group. This can be a consequence of factors that affect the preferred individual's behavior (e.g. age differences, particular internal states or predation alertness), or to intrinsic differences in their behavior (e.g. more experimented or socially dominant individuals that pay less attention to their neighbours in decision-making processes). Individual variation often causes the emergence of leader-follower dynamics (i.e. variations in individuals' probability to follow others Herbert-Read et al., 2011; Nagy et al., 2013). The mechanisms behind these collective dynamics are relevant to understand which strategies groups can adopt to transfer information and endow some members with greater weight in the collective decision-making process (Chen et al., 2016).

In the field of biology, it is accepted that groups of individuals that follow simple rules of behavior can self-organize. Nonetheless, animals are more complex than particles due to their cognition and the behavioral diversity from species to species. Moreover, they can modify their behavior substantially depending on the environmental or social context and have the further capacity to learn. As has been proven for physical systems, it is reasonable that not all animal characteristics are involved in explaining collective behavior (Camazine et al., 2001b). Yet, there is no clear delineation of which characteristics are relevant or not. This is a non-trivial problem

as what can be a qualitative good explanation is not always consistent with an empirical quantitative analysis, where differences among species, or even groups of the same species, become evident. A continuous feedback between empirical and theoretical work is indispensable to unveil the level of universality in collective phenomena and to characterize it properly (Giardina, 2008).

1.1.1 Types of collective animal behavior

Although it can be demonstrated in multiple ways, there are some iconic examples of self-organization in animal groups that exemplify different behavioral possibilities. Among them are collective decision-making, that refers to when a group of individuals collectively choose an option out of different possibilities, for example a foraging area, a new nest site or the shortest path from nest to food. Another type of collective behavior is collective motion, observed when a group of animals move together in a coherent way, like the flocking and schooling behavior of birds and fish, respectively. Finally, synchronization, a collective behavior where a group of individuals experiment periodic behaviors that become synchronized by locally exchanging information. Examples of this are fireflies light synchronization (Sarfati et al., 2021) or even clapping in an audience (Néda et al., 2000). Next, we briefly discuss the particularities of the different collective animal behaviors.

Collective decision-making

How do honeybees choose their new home? When a colony reaches a certain size, the queen leaves the hive with approximately one third of the worker bees, and clusters nearby to start the process of finding a new nest-site. A search committee involving around 5% of the bees, the scouts, fly out to inspect potential nest sites. When a scout finds a site, it assesses the site's quality and returns to the swarm. Once there, it performs communication *waggle dances* to advertise the site, and recruit others to visit their discovery and perform dances as well. The better the quality of the potential site, the longer and stronger the honeybees dance for it. Initially, the scouts visit sites randomly, but with time they are more likely to visit the sites advertised by others. Consensus on a site is reached when a site has attracted a sufficient number of bees. Some bees sense the quorum, and return to the swarm to signal the end of the

decision-making process and start the flight towards the new home. This dynamics shows the advantages of collective pooling of information, where coarse-grained information at the group level is more suited for decision-making than information held by each individual (Seeley and Visscher, 2004; Beekman and Oldroyd, 2018). Another surprising example, and one of the first animal collective behavior studies, was the way in which ants self-organize by following pheromone trails (Wilson, 1971). This is demonstrated when an ant finds a food source, and leaves a trail of pheromones at regular intervals while traveling back to the nest. Other ants, unaware of the location of the food source, encounter the trail and follow it to find the food. Once they find it, they come back to the nest and reinforce the pheromone trail by leaving traces of the substance too. It has been observed from experiments that when more than one path to the food source exists, and as pheromones evaporate over time, the path where the trail is stronger is the shortest path (Beckers et al., 1993). This way, more ants will find the trail and reinforce it on their way back to the nest. Through this mechanism, called positive feedback, multiple insects can solve collective problems such as this one (e.g. Seeley et al., 1991; Bonabeau et al., 1998).

The self-organization of honeybees and ants are just two examples of how non-verbal communication allows for collective problem-solving in social animals. Animals frequently make collective decisions. Humans make them all the time, from huge collectives like societies electing a democratic leader, to smaller groups making social decisions such as which restaurant to have dinner. Other animals agree on where to rest after traveling for a while (Harcourt and Stewart, 1994) or on which site they are going to build a new nest (Seeley and Visscher, 2004). It is apparent that collective decision making is an integral part of social animals. Decisions in a group can be made by one or a few individuals within the group, specially in small collectives, where either social dominance is imposed or some individuals have more information about the environment than the rest of the group and therefore, their opinion weighs more for that current decision making process. However, it is a more common occurrence that decision making in a group is the product of self-organization, when a group is large and communication occurs at the local level. In this scenario, each individual contributes equally to the decision outcome (Giardina, 2008).

Successful collective decisions require, most of all, correct information within the

group. In this sense, information uncertainty has a strong impact on the output of collective decision making. Typically, individuals lack pieces of the information that would help them decide between more than one option, which makes them prone to make mistakes while deciding. Consistent information sharing in collective decision making reduces individual errors and decreases the risk of making a mistake. The scale of communication also influences the way a decision is made in a group. In small groups, communication can be global, i.e. all individuals can communicate with all the others. In these cases, complex negotiating behaviors during collective decision making can take place. Large groups, however, rely on local information and their mechanisms for collective decision-making are restricted to self-organizing rules. The accuracy of decisions also depend on how many individuals contribute to the decision (Conradt, 2012). A longer quorum will increase the probability of making an optimal decision. The possibility of information pooling, leading to a more accurate decision outcome, can result in fitness advantages to consensus decision makers (Bose et al., 2017). However, a large group might take longer to reach a consensus than a smaller group. The speed restriction may create the need to adjust the number of decision makers in the group. In this way, a trade-off between speed and accuracy can be optimized.

Additionally, collective decisions require that different, and often conflicting, individual's preferences be agreed-upon. Thus, an important factor that groups have to solve to achieve consensus is the potential conflicts of interests. The presence of a conflict in the decision making process in a collective determines the exchange of information within members and the degree of cooperation during decision making. Depending on the chosen outcome, the conflict could have potential consequences in individual fitness. There are decisions that imply almost no conflict, as they benefit the group in an equal manner and every member shares the same goal. This is exemplified with the evasion maneuvers of a school of fish to escape a predator (Rosenthal et al., 2015) or honeybees choosing a new nesting site (Seeley and Visscher, 2004). However, in other situations individuals of a collective can differ in their preferred option, especially in traveling directions (Ruckstuhl, 1998) and the timing of activities (Rands et al., 2003). Even slight differences can produce conflict within the group, and to reach a resolution, some individuals will have to compromise with

other group members. Individuals' needs and physiological status can affect their level of urgency to pursue a particular option. As a result, such individuals may act more independently, prioritizing their own needs over the opinions of others, and exert a greater influence on the collective decision-making process, potentially even acting as a "temporal leader." This may come at the risk of compromising group cohesion, as proposed by (Conradt, 2012).

To study collective decision-making multiple models have been proposed that address information uncertainty and conflicting individual preferences. Models that examine information uncertainty seek to understand the way the information is transmitted within the collective, and further seek to determine how accurate a collective decision is dependent on the dissemination of the available information (Franks et al., 2009; Ward et al., 2011) Alternatively, models that include conflicts of interest deal with the adaptive mechanisms an individual could incur, which would then maximize their individual benefits within a group (Conradt and Roper, 2000; Conradt, 2012).

Collective motion

Geese collectively migrate from Canada to southern parts of North America the autumn months. While traveling, they exhibit the familiar V-shaped formation, which helps in reducing the amount of energy exerted during flight. The way we understand this spatial organization is through simple interaction rules that make neighboring individuals maintain a certain distance between them and simultaneously match their direction of motion. By varying these simple rules, it has been shown that different collective patterns can be created (Vicsek et al., 1995; Couzin et al., 2002). From disorganized groups where individuals are moving in random directions to highly organized collectives following a common direction of motion. This fascinating behavior is observed across systems of very different nature. From a wide range of biological systems, such as bacterial colonies, fish schools (Lopez et al., 2012) or human crowds (Sumpter, 2010; Silverberg et al., 2013), to non-living moving particle systems, this self-organized phenomenon occurs when many individual units interact by modifying their movements to achieve a coherent group displacement without the need of central control.

Animal collective motion presents the opportunity to link the behavioral rules that individuals follow with group-level emergent collective states. Fish schools, for example, dynamically change their shape, their speed and their direction of motion when foraging, defending or fleeing, which indicates a greater flexibility of individual interactions and rapid information transfer. Information about a threat, for example, can be transmitted through the body movements of individuals responding both directly to the threat and indirectly to the movements of its neighbors (Herbert-Read et al., 2015). Individuals survival within a collective depends mainly on the group's adaptive capacity to respond to a changing environment (Viscido et al., 2004; Calovi et al., 2015; Tunstrøm et al., 2013).

To achieve collective motion, individuals modulate their alignment and attraction toward others according to their positions and velocities. Both empirical and theoretical approximations are being developed to study the underlying mechanisms of coomunication in collective motion events (Vicsek and Zafeiris, 2012).

With the advances in image and tracking technologies a set of trajectories in space and time can be obtained for individuals within a collective (Calovi et al., 2014). From these data it is possible to compare the observations with simulated data results, and infer the underlying mechanisms of communications in collective motion events (Vicsek and Zafeiris, 2012).

In theoretical approximations, the universal mechanisms in which local rules of interaction between individual components generate emergent global patterns is being investigated with models of self-propelled particles (SPP). The simplest model is composed by moving particles at a constant speed. At each time unit, they change their orientation depending on both the average orientation of their neighbors and a noise term that randomizes the direction of motion of each particle (Vicsek et al., 1995). From this model, others have been built with varying levels of sophistication regarding the interaction rules. Examples of this include the number and properties of neighbors that the focal individual interacts with (Camperi et al., 2012) and the type of interaction depending on the particles' distance and direction (Huth and Wissel, 1992). Still, the goal of these models is to ensure that model outcomes are not dependent on some particular biological feature, but reveal universal properties of collectively moving groups.

In SPP models the interactions among particles may give rise to the collective states that are also observed in moving animals, from disorganized aggregations to directionally coordinated groups (Giardina, 2008). Additionally, these models emphasize that animal collectives on the move can be treated as a dynamical system in which multiple collective states exist . These may be dependent to how interacting features are modulated, particularly regarding repulsion, alignment and attraction among individuals (Tunstrøm et al., 2013).

Synchronization

Fireflies are one of the most outstanding examples of synchronized behavior. In some firefly species, an individual produces cyclical light flashes at night as part of his mating display. Among individuals, there are slightly different flashing frequencies. However, it has been shown that by means of local interactions, fireflies can modify their individual frequency towards the group average in order to collectively synchronize when group density is high enough (Sarfati et al., 2021). The adaptive properties, if any, of this mesmerizing behavior are still a matter of investigation (Moiseff and Copeland, 2010).

Synchronized rythmic activity is seen in many animal species (e.g. Patel et al., 2009; Kastberger et al., 2008). Experiments in the applause of a human audience have shown that after an initial phase of unsynchronized clapping, an audience is able to synchronize the applause and adopt one single frequency (Néda et al., 2000). This phenomenon was explained using a classical mathematical model for the synchronization of coupled oscillators, the kuramoto model (Kuramoto, 1975). This model demonstrates that if a number of oscillators in varying frequencies are coupled together, they will continually adjust their frequency to align closer to the average frequency. Thus, if the initial frequencies are not too different, they will oscillate synchronously with the same frequency after a certain amount of time. From this model, it has been proven in many biological systems that individuals with slightly different behavioral frequencies can synchronize by adjusting their frequency toward the average. If the frequencies difference is too large, synchronization does not occur.

1.1.2 Principles of self-organization in animal collectives

The possible existence of universality in collective behavior has encouraged researchers to create a unifying theory for the study of self-organization. Although this endeavour is proving to be very challenging, self-organizing groups share a few similarities that contribute to the emergence of collective behavior in animals (Sumpter, 2006). Many of the activities of animal collectives can be described in therms of the following principles:

- **Group variability**: It has been argued that individuals' variability in a group is particularly important in cases where multiple solutions to a problem exists. Individual variability plays a central role in the exploration of the possible options (Sumpter, 2006).

- **Positive and negative feedbacks**: Recurrent reinforcement is central to amplify particular events (Bonabeau et al., 1998). For example, local fluctuations are amplified to avoid predator attacks in a group. In this case, the reaction of one individual is copied by others which creates a new collective response. Another example is the formation of the pheromone trail in ants, as recurrent reinforcement amplifies the choice of others to follow the path. The amplification of a signal that feedbacks itself is known as positive feedback. A fast decision, however, can be compromised with respect to the accuracy of the response, which could be triggered by random fluctuations. Integrating damping or negative feedback becomes important, as it decreases the sensitivity of the collective to environmental noise, stabilizes a group response and reduces fluctuations within a collective. Typically, as a consequence, negative feedback increases the time it takes the group to make a decision. A balance between speed and accuracy is crucial for collective behavior fitness (Couzin, 2007).

- **Response thresholds**: In biological systems sometimes there exists a threshold from which a response is triggered in the group. The most typical examples are firing neurons that after reaching a voltage threshold, they activate their action potential to fire (Mensi et al., 2012). In animals, there are also many response

thresholds. For example, after a certain threshold temperature is reached bumblebees start faning to cool the nest (Weidenmüller, 2004). Another example are white-faced capuchin monkeys, that when an individual changes its direction of motion, it is only after a threshold of number of followers that the whole group collectively moves (Petit et al., 2009).

- **Leadership**: Although leadership might seem uncompatible with the idea of self-organization, it is observed sometimes as an emergent characteristic. For example, some fish in a school could have information not available for the rest of the group, such as the presence of a predator or an environmental barrier. These individuals could act as leaders of the group response as they could trigger the response of the group by changing their individual behavior (Múgica et al., 2022).

1.1.3 Transitions and adaptability in collective animal behavior

In biological systems the parameter space that we use to describe their behavior is ample. When we try to explain *why* these systems exhibit collective behavior, an evolutionary perspective appears to be the right approach. Given such parametric spaces we might ask if an optimal parameter combination exists in the proposed models that could be related to specific biological functions.

Considering this, the *criticality* hypothesis suggests that self-organized biological systems should operate close to a critical point from which different macroscopic behaviors can be easily reached, i.e. parameter combinations that place the collective near a behavioral phase transition (Munoz, 2018).

This hypothesis has been explored in different systems, such as gene-regulatory networks (Alvarez-Buylla et al., 2008) and collective behavior of animal groups (Mora and Bialek, 2011; Calovi et al., 2015), and multiple phase transitions have been described. The classical behavioral change in gregarious animals is the transition from a disordered state, with individuals moving in random directions, to an ordered state, where individuals share the same average direction of motion. Typically this can be observed in flocks, schools and herds. Another common phase transition observed in nature occurs in collective decision-making. That is, when an initially

undecided group eventually commits to a particular option, measured by the majority of individuals or other consensus definitions. For example, when honeybees choose a new nest site, discarding other options in the process.

Phase transitions are often analyzed regarding their advantages and disadvantages in the biological function of animal groups. Spontaneous symmetry breaking is an important aspect in behavioral phase transitions. It refers to the emergence of a particular global state from a collective system when multiple states are equally likely. When a group of animals make a decision for an option out of multiple equally favored ones, symmetry breaking becomes advantageous, as otherwise the group would find itself in a deadlock. Animal systems that show spontaneous symmetry breaking normally behave close to the critical point (Buhl et al., 2006). However, being near the transition represents a trade-off between being able to react fast and its robustness to keep the current behavior. In an animal collective that stays close to criticality, individuals of the group are particularly sensitive to environmental and internal noise. If an individual makes a wrong decision, misleading information could propagate through the group and cause an erroneous collective response. Thus, animal collectives are faced with a trade-off between its robustness to keep the current behavior by being able to filter out noise, and the collective's sensitivity to change by remaining highly responsive to relevant information. For this, a finer approach that goes beyond the *criticality* hypothesis has suggested the existence of mechanisms in animal groups that modulate the distance from the behavioral transition critical point according to the perceived risk and noise of the environment.

How animal groups place themselves at a particular distance from criticality is still a matter of research (Hidalgo et al., 2014; Klamser and Romanczuk, 2021). However, experimental works have started yielding some hypotheses. Macroscopic properties have been related with particular variables that seem to control the group's distance to criticality, which include the group size, the group density and the group heterogeneity in the individuals' behavior (Moretti and Muñoz, 2013; Attanasi et al., 2014; Sosna et al., 2019).

1.2 Using robotics to study animal collectives: an emphasis on social insects

1.2.1 Social insects as an inspiration for swarm robotics

The field of swarm robotics has drawn inspiration from studies of the collective behavior of animals, such as fish schools, bird flocks and, particularly, social insects (Garnier, 2011).

Insects form one of the most successful phylum on earth, with one million discovered species ouf of an estimated 5.5 million overall. They are adapted to a multitude of environments with different types of locomotion (e.g. crawling,flying, swimming), complex visual systems, navigation strategies and collective behaviors (Stork, 2018).

The colonies of social insects are emblematic examples of implementing self-organization to achieve an impressive array of behaviors, such as maintenance of the colony (Spivak and Danka, 2021), complex nest architectures (Sane et al., 2020), defense against invasions and sophisticated divisions of labor(Abbot, 2022).

The fascinating collective behaviors exhibited by insects are the result of local interactions among individuals with limited cognitive abilities. Despite these limitations, insects have the capability to learn from each other in various aspects of societal living, such as foraging options, predation threats, and potential nest sites. This demonstrates the efficiency, flexibility, and robustness of insect collective intelligence (Leadbeater and Chittka, 2007).

Swarm robotics involves the utilization of multiple robots working to achieve a common goal through collaboration. The key objectives of swarm robotics include the exploration of an environment as a group, locating targets, transporting objects through cooperation, and assigning tasks without external supervision (Garnier, 2011). For these purposes, the primary objective of swarm robotics is to build systems that are robust yet flexible to changing environments, so they can handle unexpected circunstances (Dorigo et al., 2021), i.e. they are able to "expect the unexpected". The basis of this idea is that collective behaviors will emerge from the local interactions of swarms of simple robots (Sharkey, 2006).

It is not surprising that this field has drawn inspiration from social insects in an effort to compose collectives of robots. Social insects are an archetypical example example that self-organization provide adaptive solutions to complex problems (Dorigo et al., 2021). Seeking to imitate insect intelligence can achieve higher autonomy levels, even with limited processing capabilities in small robots (De Croon et al., 2022; Rubenstein et al., 2012).

1.2.2 Swarm robotics as a tool to study collective behavior

As previously mentioned, social insects exhibit many of the behavioral characteristics from which the field of swarm robotics benefits. The inverse notion, that robotics can be useful in the study of insects societies and of biological systems in general, has recently been explored in more depth (e.g.Arvin et al., 2011; Halloy et al., 2007).

Experimentation with swarm robots to address animal behavioral questions may be consideed a slow and inconvenient alternative to theoretical modeling and numerical simulations (Garnier, 2011). Nonetheless, robot systems address real world scenarios in a way that neither mathematical models nor numerical simulations can. The study of animal collective behavior through the use of robotics highlights the importance of considering the "situatedness" of individuals, defined as the extent to which individuals are embedded within an environment that they can both sense and modify. This approach aligns more closely with animal empirical studies and emphasizes the connection between the individual and its environment (Varela and Bourgine, 1992).

Empirical studies face a challenge in determining the causality behind an observed behavior. It is difficult to establish whether an individual's behavior is a result of a particular event, especially when conducting field studies where control over variables is nearly impossible (Mitri et al., 2013). On the other end of the spectrum, abstract mathematical models can also be implemented for the study of animal collective behavior. These models make predictions by reducing the collective system to its core elements and examining the influence of a limited number of parameters on the system's dynamics (Bose et al., 2017). The level of abstraction in these models, though, simplify relevant aspects of individuals' inner status and of

the environment in which they develop. To include such factors, self-propelled particles (SPP) and agent-based models were developed. These computational models allow simulations of the behavior and interactions of autonomous agents as a way to understand the global behavior of a system (Mitri et al., 2013). Nonetheless, individual agents are described probabilistically, which reduces the individual variation that exists in nature, as an animal's behavior is a product of complex combinations between perceptual, cognitive and locomotor activities (Garnier, 2011).

In the last two decades, researchers have resorted to swarm robotics as a means to study the emergence of complex behaviors in natural systems (Dorigo et al., 2021). A physical robot refers to a *machine that is able to interact physically with its environment and perform some sequence of behaviors, either autonomously or by remote control* (Krause et al., 2011).

The study of collective behavior through the lens of robotics is similar to that of agent-based models, where individual robots interact with each other. However, the fact that physical robots execute their behavior within the constraints of a real-world environment, obviates the need to make assumptions regarding its properties, such as differences in perception or spatial limitations (Mitri et al., 2013). In other words, swarm robotics *situatedness* implicitly include the laws of physics, which dictates the limitations imposed on the individual robots' mobility, perception, and information transfer within the collective, as well as the impact of environmental noise. In systems where the physical environment and spatial considerations play a decisive role in determining collective behavior, the use of robotic systems becomes an indispensable tool for research. For instance, the agent-based modeling of aggregation behavior in the German cockroach (Blatella germanica) resulted in larger clusters than the actual animal groups (Jeanson et al., 2005). It was not until researchers resorted to physical robots that it became apparent that the extended body of the robots reduced individuals' ability to sense others in the cluster, resulting in smaller aggregates more akin to those observed in real cockroaches (Garnier et al., 2008). An example of how physical interactions can influence the movements of the individuals in a group comes from the study of foraging efficiency in ant-like robots programmed to collect objects scattered in an enclosed space. The study revealed that gathering efficiency was reduced in larger groups due to increased interference

among robots, compared to smaller groups (Krieger et al., 2000). This finding offers a apotential explanation for the phenomenon of decreased "per capita" productivity with increasing group size observed in various social insects (Kramer et al., 2014).

Moreover, swarming robot models are more likely to result in unexpected outcomes due to the real-world properties that may not be considered in more abstract approaches (Trianni et al., 2016). Furthermore, the emergence of global collective behaviors from the actions and interactions of swarming robots in response to real-life stimuli provides evidence for the feasibility of the mechanisms proposed to underlie collective output. For instance, robots can be tested in conditions that closely resemble the environment experienced by certain social insects, and the behavior of the swarm of robots can be compared to that of the target organisms under similar conditions (Sharkey, 2006).

Studies about collective behavior have covered different ranges of situatedness by incorporating robots in simulated and physical form, and experimenting with mixed models where physical robots directly interact with animals. Computational models of simulated robots are crucial in swarm robotics, serving as a complementary tool to physical robots. These simulations are designed to mimic the behavior and characteristics of the corresponding physical robots and their environment. The simulated robots are not simple points as in most agent-based models, but rather have an extended body that occupies space in the world and sensors with limited range and accuracy that gather information about their environment. Additionally, they move within a defined space with limited resources. The fact that the simulations mirror the behaviors of physical robots forces researchers to consider the limitations in movement, perception and resources of the real world when conducting experiments (Mitri et al., 2013). Simulations are an important tool in research as they allow for multiple experiments to be conducted with numerous individuals, even more than physically available, in a fraction of the time required for physical experiments. Before expanding physical robot research, simulations must be calibrated to be a reliable substitute for physical robots (Mondada et al., 2004). For example, in the context of collective behavior, simulations must adjust certain parameters such as interaction range or perception limitations due to environmental factors so that they can accurately reproduce both individual and collective behaviors.This way,

valid conclusions can be drawn with increasing statistics.

However, one of the major challenges in real-world simulations is known as the simulation-reality gap, defined as the possibly subtle but important discrepancy between the real-world and simulations. This difference can have a significant impact on the performance of physical robots in real-world applications of swarm robotics. The ideal swarm robotics simulator should minimize these discrepancies, although they cannot be completely eliminated (Bose et al., 2017).

Robotics to emulate and study insects behavior

In recent years, the use of physical robots in the study of collective behavior has gained significant attention due to the unique advantages that they offer in comparison to traditional methods. Insects and robots share some similarities. Insects appear to follow a set of predetermined behaviors in response to environmental stimuli, and their interactions with each other can seem fixed and repetitive, leading to a robotic-like appearance.

Insect intelligence is characterized by its parsimony, defined in this context as the efficient use of minimal resources to achieve successful behavior in complex environments. This parsimony is dependent not only on the capacity of the brain, but on insect embodied cognition, that is, an insect's body and sensory apparatus that expand the perception and interaction of each individual. In other words, insect intelligence builds on the capacity to interact with the world through embodiment, simplifying the cognitive effort of the brain. Through this lens, working with robots to study insects behavior is not only ideal, but perhaps even crucial to replicate both the situatedness and the sensorial characteristics that contribute to their intelligence in real world environments (De Croon et al., 2022). The swarm intelligence of insects further allows for parsimonious solutions, offering robustness, scalability, and flexibility that can be also replicated in swarming robots. Robotics research has much to learn from insects and their adaptive response to the environment, and as our understanding deepens, so should our ability to emulate them (Sharkey, 2006).

For these reasons, researchers have created robots to artificially replicate insects' behavior and to increase our understanding of these life forms (Rubenstein et al., 2012; Ma et al., 2013; Phan et al., 2019; Yang et al., 2020, Figure 1.2). To be considered

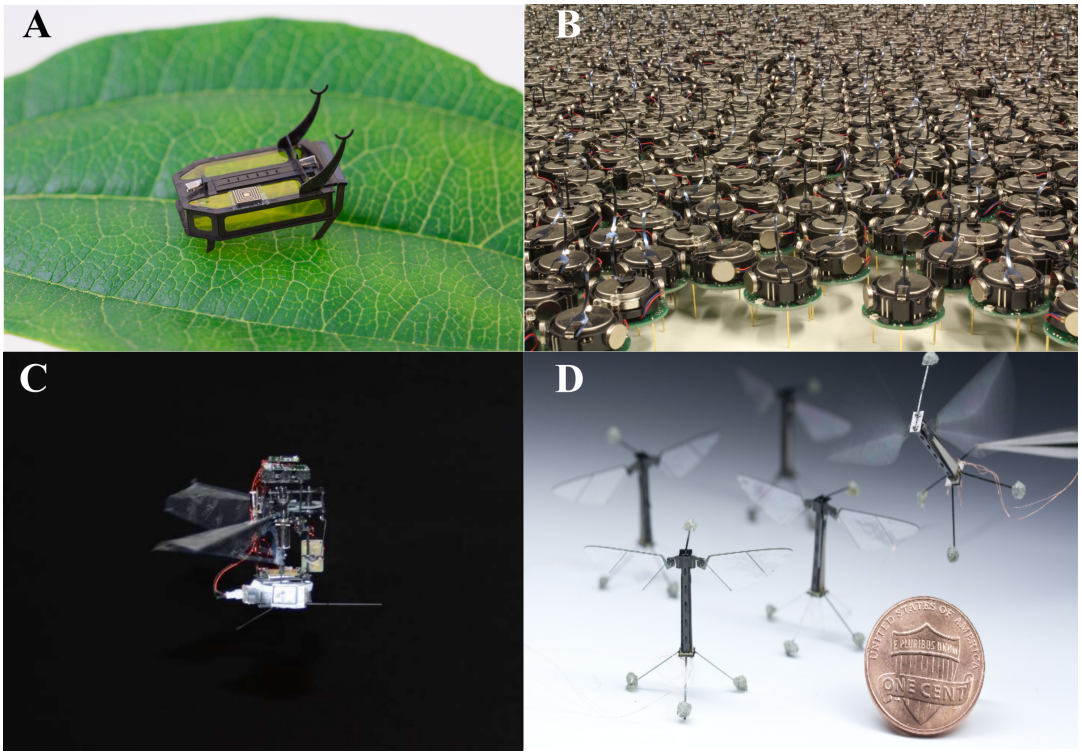


Figure 1.2: Robots inspired by insects' bodies and behavior. **A**: RoBeetle, an autonomous crawling robot that emulates muscles and metabolism in beetles (citeyan2020; image borrowed from the authors paper). **B**: Kilobots, swarming robots that can perform collective tasks (Rubenstein et al., 2012); image from wikimedia with licence CC BY-SA 4.0. **C**: KUBeetle-S, a flying robot that can perform stable flight (Phan et al., 2019; image borrowed from the authors' paper). **D**: Fly-like robot with flapping wings (Ma et al., 2013; image borrowed from the authors' paper)

effective, robots have to exist in a physical form and not just in a computer simulation, and interact within a real-world environment. It is through these interplays that both their capabilities and limitations are studied and understood (Webb, 2001).

The self-organization of insects has been studied particularly in the field of robotics. For example, the ability for self-organization without the need for communication or memory, only using simple oscillatory processes, was tested in groups of physical robots. The experiments resulted in synchronized light-emission patterns among groups of physical robots, which is thought to be analogous to synchronized firefly light production (Wischmann et al., 2006). Other studies have been focused on understanding the evolution of communicative behavior in robots. This question has been challenging to address through other experimental methods due to the long time scales often associated with the evolution of animal communication and the lack of a fossil record. However, experiments using both physical and simulated robots have been instrumental in showing that efficient self-organization processes

can occur even with little sensory information. For example, robots following simple rules have been observed to self-organize into groups performing different tasks, illustrating that division of labor can take place simply due to differences in local perception (Mitri et al., 2013).

These studies have expanded our understanding of self-organization in insects, and have demonstrated the potential for robots to provide more convincing demonstrations of these phenomena than is possible with just software.

Overall, the use of physical robots in the study of collective behavior is a rapidly evolving field that offers immense potential for increasing our understanding of the underlying mechanisms of social behavior. By studying the similarities and differences between insects and swarming robots, researchers can gain insight into the mechanisms underlying swarm intelligence and use this knowledge to design more intelligent and adaptive systems.

1.3 Scope of the thesis

The primary goal of this thesis is to explore the underlying mechanisms from which complex collective behavior can emerge and adapt. We concentrate our analysis in the collective motion of schooling fish and consensus reaching in swarming robots in a model of honeybees nest site selection.

Our work is divided in two parts. The first part concentrates on the collective motion of black neon tetra (*Hyphessobrycon herbertaxelrodi*), a small schooling fish that inhabits in freshwaters from southern Mexico to Argentina.

In Chapter 2 we describe the characteristics of shoaling behavior in fish and the main properties that are measured to further comprehend their collective motion. We then explain our experimental setup to record groups of 20 and 40 black neon tetra.

In Chapter 3 we discuss how animal tracking methodologies are typically structured, and describe the design of a custom tracking software we developed to extract the individuals' (x, y) positions in time for further analysis of their collective behavior. To verify our results, we categorize the individual features of the fish and compare them with reported behavioral traits of our fish collective and other closely related species.

Chapter 4 is dedicated to describing existing models of self-propelled particles (SPP) which study collective motion. In our research, we focus on a classical simple SPP model, the Vicsek model. We explore the behavior of the system's order-disorder phase transition as a function of the system's parameters, including unreported transitions as a function of particles' speed.

Our first empirical and theoretical analysis of collective motion is carried out in Chapter 5. Here, we study the relationship between swimming speed and collective ordering in black neon tetras, and we imitate the speed behavior observed by modifying the Vicsek model to include a periodic oscillation of the particles' speed.

In Chapter 6 we investigate sudden directional rearrangments observed in our fish collective. We analyze the role of effective leadership in initiating these behavioral cascades and replicate the behavior by modifying the Vicsek model to include a global leader.

The second part of the thesis is dedicated to the analysis of a decentralized model of the nest site selection in honeybees and its implementation in a more realistic scenario by utilizing small swarming robots.

In Chapter 7 we describe the characteristics of the swarming robots, the kilobots, and categorize their individual speed and message reception. We also engage the kilboots simulator, kilombo, to observe the motion of a kilobot collective moving as persistent random walkers within a circular arena.

In Chapter 8 we indicate different models for collective decision-making and concentrate on existing models of honeybees nest site selection. We elaborate on the first model to integrate the interaction between independent decision-making and imitation, and further categorize the behavior of the system in a mean-field approach by exploring its parameter space.

In Chapter 9 we implement the model within our kilobot system. We describe our experimental design and setup, carry out the empirical experiments and compare our results with mean-field approximations and quenched configurations. We further explain our results by means of percolation theory and network methodologies.

In the final Chapter, Chapter 10, we summarize our findings and provide insights that can guide future research.

Part 1

Collective motion of schooling fish

2 Empirical analysis of shoaling fish

2.1 Introduction

Social aggregation is a common feature in at least half of all fish species throughout their life cycle (Pitcher, 1998). It is probable that aggregation serves as an adaptive mechanism which helps fish respond within a variety of environmental contexts and necessities. The primary advantages of grouping are expressed in foraging and anti-predation maneuvers. Foraging together allows for the fish to find food more rapidly (Pitcher et al., 1982) and to spend more time eating with a decreased need for vigilance (Magurran et al., 1985). Anti-predatory advantages include a reduced individual risk of being taken by a predator as group size increases (Foster and Treherne, 1981). This is due, among other phenomena, to the perceptual or cognitive overloading that predators experience when confronted with a large group of similar-looking prey, rendering them unable to choose between multiple items of prey (Krakauer, 1995). Also, as a group the fish are more rapidly able to detect a predator (Ward et al., 2011) and their capacity to counter an attack by means of synchronized evasion maneuvers increases (Pitcher, 1983). Other advantages for group members include access to potential mates (Krause et al., 2022). Additionally, there is evidence of hydrodynamic benefit to group movements, as fish save energy and decrease oxygen consumption by taking advantage of the flows and vortices created by the fish in front of them (Weihs, 1973; Liao et al., 2003). Living in a group can also yield disadvantages, such as individuals' detection by a predator is more likely to occur in large groups of fish than in small groups and solitary individuals, and both competition for food and contagious ectoparasites exposure also increase

with increasing group size (Ruxton, 2012). Therefore, fish constantly perform a risk-balancing trade-off and decide to stay with a group, go off alone or join other groups of fish (Lopez et al., 2012).

The adaptive nature of aggregation results in the vast repertoire of collective behaviors that we observe in nature. Collective motion patterns are thought to emerge from simple behavioral rules of individual motion, and from the interaction between individuals and their environment (Parrish and Edelstein-Keshet, 1999; Herbert-Read et al., 2011). The degree of aggregation in a group is characterized by its cohesion, defined by the spatial distribution of individuals. When the individuals present a significant degree of cohesion the group is defined as a shoal. Consequently, shoaling behavior is often described with spatial parameters such as distance between individuals, surface area or volume measurements.

Moreover, shoals can be dynamically structured with different degrees of motion synchronization. Shoals that aggregate but are locally and globally disordered are called swarms. A shoal in which fish have a tendency to adopt similar orientations and match their swimming speeds is called a school (Delcourt and Poncin, 2012). In a school, fish can be synchronized both locally and globally, or they can show a milling behavior, which is characterized by a high degree of alignment with local neighbors but globally the school forms a rotating mill (Tunstrøm et al., 2013). An example of these dynamic states is shown in Figure 2.1. Schooling behavior is measured not by spatial parameters, but by dynamic parameters involving vectors and angles. With these parameters, we can determine a degree of school behaviors for a given social group at a given time (Delcourt and Poncin, 2012).

Schooling may have appeared in a very early stage of vertebrate evolution (Vicsek and Zafeiris, 2012). It is estimated that at least 50% of fish species school as juveniles, and that approximately 25% of species school throughout their lives (Lopez et al., 2012). Depending on the species, fish can spend more or less time schooling as a response to different environmental contexts. For example, the synchronization of fish movements decrease predation risk as it is thought to accelerate the transmission of information about a potential predator (Delcourt and Poncin, 2012).

Schooling is relevant to many fields in biology, including ecology, ethology and

Figure 2.1: Swarming, schooling and milling behavior in fish shoals
Swarm of maldives surgeon fish (left), school of common bluestripe snappers (right up) and mill of barracudas (right bottom). Credits: Robin Hughes (licence CC BY-SA 2.0), Uxbona (licence CC BY 3.0) and Bernard Dupont (licence CC BY-SA 2.0).

neurosciences. Moreover, understanding schooling behavior has a practical implication for the fishing industry, as fisheries can adopt better fishing techniques. With these interests in mind, multiple studies in the last decades have categorized schooling phenomena to address questions regarding the dynamic patterns exhibited in natural and experimental conditions. Observations in natural environments have primarily been conducted by video recordings from the air and underwater (Doran et al., 2022), and by sonar techniques that allow to estimate the size and density of the schools (Gerlotto et al., 2000). To increase control of the enviornmental variables influencing groups, experimental observations are carried out in controlled laboratory setups and individuals' behavior is recorded by zenithal or lateral cameras to analyze their motion either in real time or deferred in time.

It is clear that schooling is a biologically relevant feature. The underlying mechanisms from which collective motion emerges are central for schooling to be adaptively successful (Parrish et al., 2002). Considering this, a dominant question in collective motion research is evaluating how individual behavior and interactions within the group result in the large-scale patterns we observe in nature. It is through the analysis of these underlying properties at local and global scales that it can be determined which collective behaviors are adopted by an animal aggregation (Delcourt and Poncin, 2012).

2.2 Shoaling characterization

The properties of animal collective motion resemble other physical phenomena like ferro-magnetic particles or molecular spatial arrangements, which has rasied the interest of the statistical physics community and led to the application of statistical physics tools to gain further insight into the structure of animal collective motion.

Both group cohesion and motion synchronization are analyzed to study collective motion in fish. In order to characterize group cohesion, descriptive parameters involving the euclidean distance between individuals are obtained. Synchronization is commonly measured by an individual' speeds angle differences and the degree of polarization, which refers to the degree of direction alignment among individuals.

The requisite information for these determinations are the x, y coordinates of each individual in time. From this, group cohesion can be described by the distance of a focal fish i to its nearest neighbor nn_i at time t,

$$NND_i(t) = \sqrt{(x_i(t) - x_{nn_i}(t))^2 + (y_i(t) - y_{nn_i}(t))^2}. \tag{2.1}$$

A more global descriptor of group cohesion can be obtained by calculating the average distance of all pairs of individuals in a group of size N at time t,

$$IID(t) = \frac{2}{N(N-1)} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} D_{i,j}(t), \tag{2.2}$$

where $D_{i,j}(t)$ is the distance of fish i to fish j at time t. This descriptor can be corrected by body size to obtain a relative distance in body lengths,

$$IID_{BL}(t) = \frac{IID(t)}{BL}. \tag{2.3}$$

Still, differences in body size within the group can bias the cohesion quantification.

Another common parameter that has been used to provide information about the degree of group cohesion is the Clark-Evans Index (Clark and Evans, 1954), a descriptor commonly used in ecology which consists of comparing the NND obtained from the experimental observations with the theoretical NND^* if the distribution of

individuals in space was random. The group mean distance to the nearest neighbor $\overline{NND}$ at time t is calculated as the sum of all measured nearest neighbor distances divided by the group size,

$$\overline{NND}(t) = \sum_{i=1}^{N} \frac{NND_i(t)}{N}. \tag{2.4}$$

The mean distance expected in a random spatial distribution for an infinite population of a density ρ is calculated as

$$\overline{NND^*} = \frac{1}{2\sqrt{\rho}}. \tag{2.5}$$

Then the Clark-Evans Index R at time t is,

$$R(t) = \frac{\overline{NND(t)}}{\overline{NND^*}}. \tag{2.6}$$

This index ranges from $0 \leq R \leq 2.1491$. When $R = 0$ there is a limit situation of complete aggregation, when $R = 2.1491$ the individuals distribution is perfectly homogeneous and when $R = 1$ the individuals are distributed randomly in space.

However, this quantification of cohesion does not consider the volume of individuals, which could differ enormously. To account for this, one can measure the packing fraction, which refers to the ratio of the individuals' exclusion volume on the group volume. Small values of the packing fraction correspond to gas-like systems, while for large values the spatial structure will resemble more a liquid or a solid.

To categorize schooling behavior the main descriptor is the velocity vector, from which headings, speed and polarization can be analyzed.

Individual speed at time t is obtained from the velocity vector $\vec{v}_i(t)$,

$$v_i(t) = |\vec{v}_i(t)| \equiv [\vec{v}_i(t) \cdot \vec{v}_i(t)]^{1/2}. \tag{2.7}$$

Speed synchronization can be studied comparing fish i speed with the speed of fish j, and correlating their values through time. Normally, correlation coefficients closer to 1 indicate that the fish are synchronized in their speed, while values closer to 0 demonstrate nonsynchronization.

Local synchronization in the direction of motion can be analyzed by the *heading difference* between pairs of individuals, $h_{i,j}(t)$, defined by the angle formed by the velocity vectors $\vec{v}_t(t)$ and $\vec{v}_j(t)$, and computed as

$$h_{i,j}(t) = \arctan\left\{\frac{\|\vec{v}_i(t) \times \vec{v}_j(t)\|}{\|\vec{v}_i(t) \cdot \vec{v}_j(t)\|}\right\}, \tag{2.8}$$

where $\times$ represents the vectorial product and $\|\cdot\|$ the vector modulus. $h_{i,j}(t)$ takes values in the interval $[-\pi, \pi]$. Figure 2.2 shows a representation of cohesion and motion synchronization measurements in two fish.

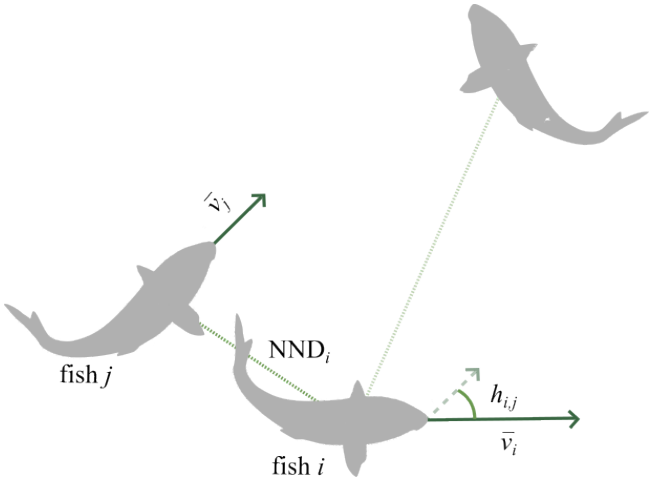


Figure 2.2: Local measurements of shoaling and schooling tendencies in two fish.

To estimate the synchronization in the direction of motion of the shoal at time t, i.e. the degree of polarization, the order parameter ϕ is commonly used (Vicsek et al., 1995; Ginelli, 2016),

$$\phi(t) = \frac{1}{N}\left|\sum_{i=1}^{N}\frac{\vec{v}_i(t)}{v_i(t)}\right|, \tag{2.9}$$

In a perfectly ordered school, in which all fish travel in the same direction ϕ takes its maximum value 1. In a disordered swarm, with all fish traveling in randomly chosen directions, a low value is expected, scaling as $\phi \simeq N^{-1/2}$ (Ginelli, 2016).

Polarization can also be measured locally considering only a subgroup of neighboring individuals of the collective, ϕ_l. In this case, the fact that local polarization tends towards 1 does not mean that the school is globally polarized, as shown in Figure 2.3 where different individual headings configurations are exemplified with their corresponding local and global polarization values.

Another order parameter that yields information about the structure of the school is the rotation order parameter ϕ_r, which describes the degree of rotation of the collective around its center of mass. It is defined by the mean angular momentum,

$$\phi_r(t) = \frac{1}{N}\left|\sum_{i=1}^{N}\frac{\vec{v}_i(t)}{v_i(t)} \times \vec{R}_i^{CM}(t)\right|, \tag{2.10}$$

where $\vec{R}_i^{CM}(t)$ is the vector pointing from the center of mass of the shoal towards fish i at time t. The rotation order parameter also ranges from 0, which indicates that the shoal is not rotating, to 1, which implies that the collective is experiencing a strong rotation. Identifying the degree of group rotation allows one to recognize milling behavior, and differentiate it from swarming when global polarization is low. Figure 2.3 exemplifies schooling, milling and swarming behaviors with their respective ϕ, ϕ_l and ϕ_r values.

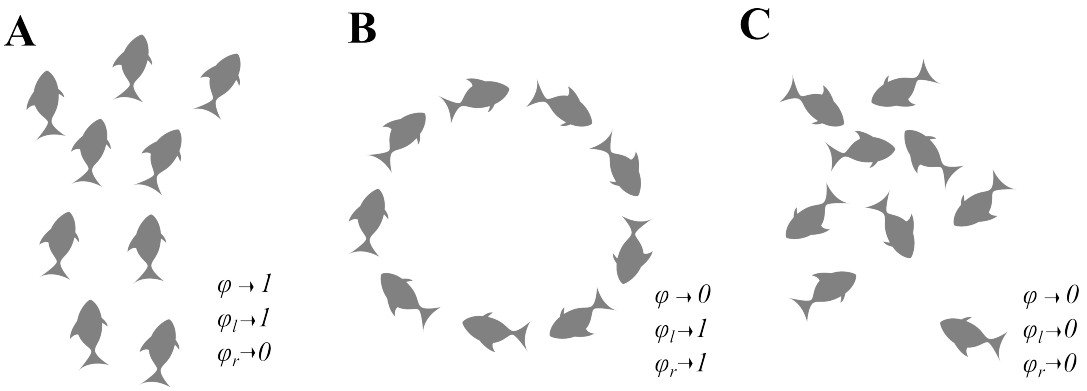


Figure 2.3: Examples of fish shoals dynamical structures. **A**: Schooling behavior with paralel polarized fish. **B**: Milling behavior. **C**: Swarming behavior.

From another research perspective, experimental studies in fish collectives have shown that the individuals' arrangement in space during collective motion could be influenced by social preferences in some species (Barber and Wright, 2001; Lee-Jenkins and Godin, 2013). For example, guppies (*Poecilia reticulata*) prefer to shoal with individuals that they are familiar with (Griffiths and Magurran, 1999) and with individuals of the same size and sex (Croft et al., 2004). To investigate the underlying social patterns and how they could impact the spatial positions of individuals in a shoal, there have been recent attempts to integrate social information in the analysis of collective motion in fish by means of network theory (Croft et al., 2004).

Network analysis provides a framework to express the relationships and interactions between individuals (Krause et al., 2009; Sih et al., 2009). From this, the structure of the resulting communication and social networks can be analyzed with the purpose of both understanding how social preferences affect the cohesion and synchronization in collective motion, and how collective motion shapes social preferences in short or long terms. From the resulting networks, relevant measures can be applied such as the distribution of contact durations, i.e. how long two fish swim together within an interaction radius (Borrel et al., 2008; Murakami et al., 2015), or the delay between the direction of motion of pairs of individuals from which a leader-follower network can be established (Nagy et al., 2010).

In the next Section we will describe our study subjects used to analyze collective motion in fish aggregates. We will study the possible underlying interactions found in our groups of fish with theoretical approaches by means of numerical simulations.

2.3 Study subjects

For our collective motion analysis we worked with black neon tetra (*Hyphessobrycon herbertaxelrodi*), a small freshwater fish of adult mean body size 2.5cm shown in Figure 2.4. The black neon tetra is one of the most popular ornamental fish species, characterized by a mild temperament. There are not many studies about the behavior and ecology of this particular species. The genus Hyephesobrycon is one of the biggest genera of the Characidae family with around 130 representative species distributed in the neotropical region from southern Mexico to Argentina (Carvalho and Langeani, 2013; Lima et al., 2014; Teixeira et al., 2016). Particularly, black neon tetras live in groups in streams and lakes. Their diet is omnivorous, consisting of worms, crustaceans and plants.

Recent laboratory research on the collective behavior of this species has shown that black neon tetra occupy mostly the upper water layers of the tank even with access to an increased water depth. They swim mostly in schools with consistently high global polarization of $\langle \phi \rangle = 0.88, 95\%$ with a Confidence Interval of $CI[0.88, 0.85]$ for groups of different sizes (from 2 to 20). They tend to remain cohesive, even

Figure 2.4: Left: individual of black neon tetra. Right: school of black neon tetras. Credits: Brian Gratwicke (licence CC BY 2.0) and Tifr (licence CC BY-SA 3.0)

at the expense of a temporary loss of polarization, showing small variations in interindividual distances, $IID = 13.56$ cm, 95% $CI[12.36, 14.74]$, and group density $\rho_g = 0.052$ fish/cm^2, 95% $CI[0.045, 0.060]$. Nevertheless, measures of heading differences and turning angles show that fish tend to turn away from neighbors if the distance between them is less than a body length to avoid excesive proximity and collisions The characterized motion of black neon tetra indicates that this species is suitable for studying collective motion, as it is a small species that swims in groups with a stable coordinated motion. In a more practical sense, this species distributes horizontally in the water column, which reduces measurement error when recording in $2D$ from a zenithal point of view (Gimeno et al., 2016; Quera et al., 2019).

2.4 Experimental setup

Experiments were performed at the Quantitative Psychology Unit of the Institute of Neurosciences, University of Barcelona. To later obtain the individuals' x, y positions in time and perform collective motion analysis, we recorded groups of 20 and 40 black neon tetras swimming freely in a $100 \times 93 \times 40$cm rectangular experimental tank, using a CCD (uEye UI-1640LE) fixed camera mounted 2.2m from the center of the tank. The signal was recorded with the software uEye Cockpit (IDS GmbH, 2010) with 20 frames per second, 8 bits of gray levels and a resolution of 1072×1004 pixels per frame. Light was adjusted from three bulb lamps and two fluorescent lamps placed around the tank to illuminate the opposite side and obtain homogeneous, indirect, low-intensity lighting, as shown in Figure 2.5.

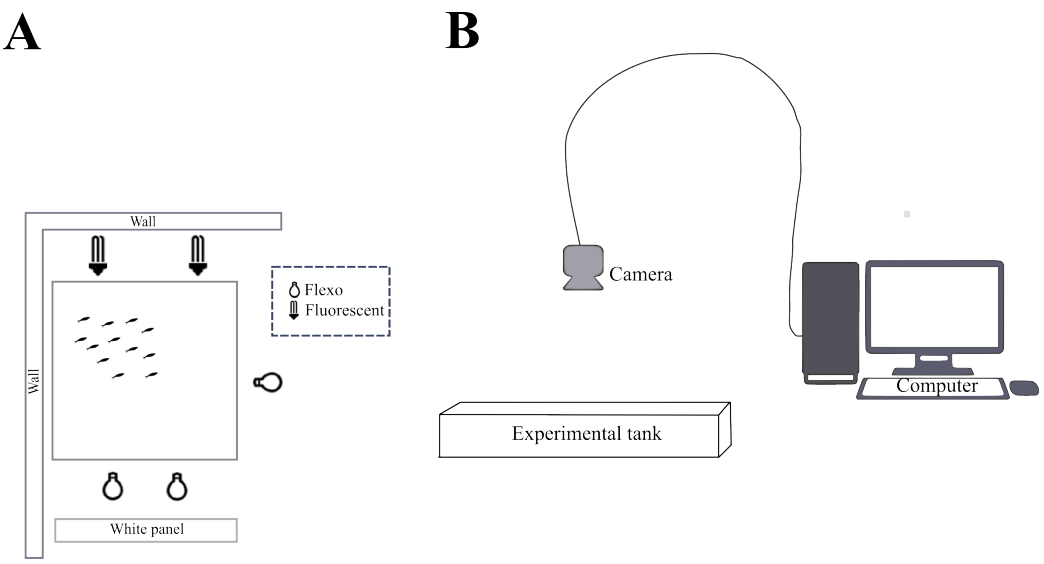


Figure 2.5: Experimental setup for the analysis of collective motion in black neon tetras. **A**: View from the top of experimental setup. **B**: Lateral view of experimental tank and senital camera, connected to the computer that records the videos.

For clarity of recording and analysis, we worked with a water column with a depth of 5 *cm* to avoid fish distributing in more than one vertical layer. Three independent replicates of 12000 frames (coded A, B and C) were recorded for each experimental condition, along with 250 frames of the empty tank for background subtraction. Between recordings fish were maintained in aquariums of identical size $40 \times 43 \times 30$ cm with light/dark regular photo-periods, at a temperature of $25^\circ \pm 2^\circ$, pH of $7.8 - 8$, total hardness of $10 - 14$ ppm, nitrate levels below 100 mg/L and nitrite levels of 0 mg/L. For each experimental trial, either 20 or 40 randomly chosen fish were moved from the aquarium to the experimental tank. First, to acclimatize the fish to the temperature in the tank the group was transferred for five minutes to a container with water from the experimental tank and from the aquarium. The group was then transferred into the tank and left to habituate to the new environment for five minutes prior to starting the recording. After recording each independent replicate, fish were captured with a net and transferred back to the aquarium along with water from the experimental tank to aid in the reacclimation. The recording and the posterior data analysis were blind with respect to the subjects, given that individual identification between recording events was not feasible. Video files were saved in AVI format for the posterior processing of individual trajectories.

3 Tracking animals on the move: Software design and implementation

3.1 Introduction

To study the individual interactions between animals from which collective phenomena emerge greatly depends upon our ability to track the individuals' behavior in space and time. The quantitative analysis of animal behavior is crucial in a variety fields, such as ecology, medicine, neurology and evolutionary research, where multiple individuals need to be tracked simultaneously (e.g. Fernando and Lande, 2000; Cachat et al., 2011; Bownik and Wlodkowic, 2021). Direct observations have lower resolution and often result in a small sample of individuals and fewer data points (Dell et al., 2014). Fortunately, the advances in technology have made large amounts of high precision behavioral data collection possible with individual tracking. This has augmented the wide interest in animal behavior with a diversity of technological solutions that address the multiple scenarios where animals are to be quantified, from tracking devices attached directly to the individuals in their natural habitat (e.g. Daniel Kissling et al., 2014), to image-based tracking in laboratory settings where software tools calculate individuals positions in time (e.g. Pérez-Escudero et al., 2014).

Organisms move rapidly when interacting with each other and their environment. To capture their motion, behavioral studies first need to adjust video frame rate, field of view and image resolution with the required temporal and spatial resolution. Software and algorithms that can track individuals positions are critical

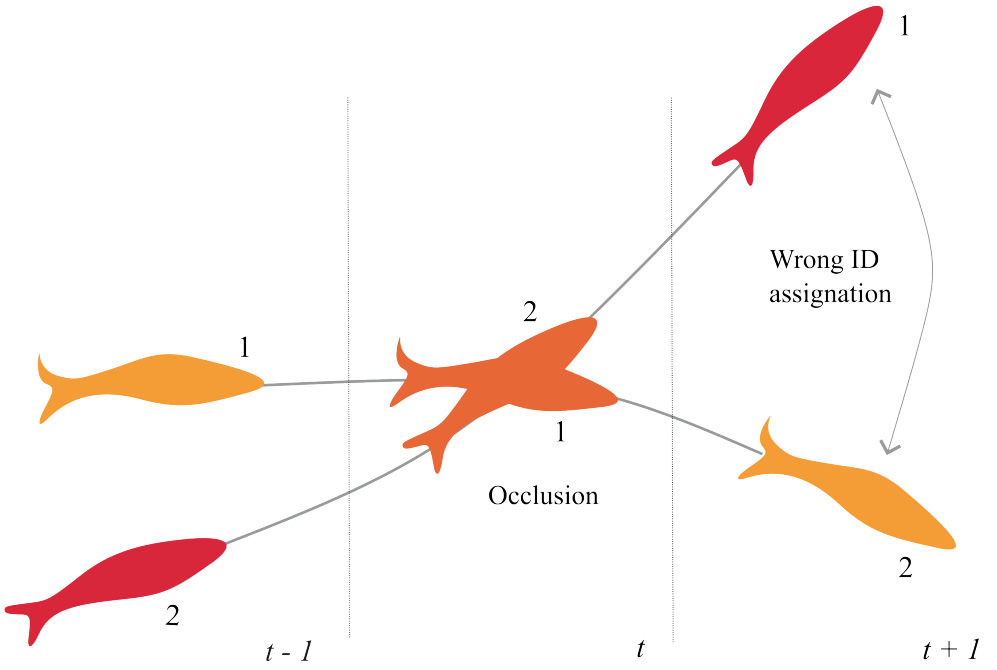


Figure 3.1: Example of an occlusion between two fish.

for efficient research (Panadeiro et al., 2021). This often implies a high amount of image data to process and analyze, which can be very labor intensive, especially if manual tracking is used. In addition, tracking individuals manually introduces the possibility of human error, which can reduce the accuracy of the results. Therefore, automatic approaches for animal tracking are needed. However, most approaches, as it will be discussed later, perform automatic tracking using human intervention for specific conflict solving.

One of the main problems of tracking software is wrong individual's identification due to occlusions, observed when individuals move really close to each other or even superimpose. When this happens, they can appear to be a single individual for a certain amount of time (Dolado et al., 2015). Occlusions can impact the tracking accuracy by mixing the identities and breaking the trajectories. In Figure 3.1 an example of an occlusion between two individuals is shown. This may have unfortunate consequences, as identity errors can propagate throughout the remaining sequence.

Addressing the issue of occlusions is an important feature of tracking software.

While there are methods capable of effectively tracking one or few individuals, preserving the identity of multiple individuals after an occlusion is still a challenging problem (Rodriguez et al., 2018). Automated techniques able to accurately track multiple organisms and preserve their identities still need to be developed.

To address the occlusion problem some technical approaches are:

- Tagging organisms with a visual marker. This technique can be physically invasive and it is not applicable in general, for example, for organisms with small body size.

- Using several cameras. This approach adds complexity to the experimental setup, as different images need to be rendered or processed together. Also, it adds computational cost to the system because of the process to integrate the different images.

- Specific models of the animal body. These methods can only be applied for animals geometrically compatible with the used model. Tracking software using this approach often relies on specific shapes.

- Sophisticated contour representations or fingerprinting (characterizing individuals by specific bodily identifiers; e.g. variations in size). These approaches greatly enhance the usefulness of image-based tracking allowing tracking of differently sized and shaped organisms (Dell et al., 2014). However, it requires videos with high-resolution.

- Image pattern recognition. This approach either uses matching texture maps or convolutional neural networks. It is computationally and memory expensive and requires access to past and future frames. Adittionally, it requires videos with high temporal and spatial resolution (Rodriguez et al., 2018).

- Customized automated algorithms which predict identity based on the relative speed and direction of movement can reduce mistakes, but error propagation is still unavoidable because of the stochastic behavior of the organisms. These software rely on human expertise to solve occlusions. However, differentiating individuals during an occlusion can be partly overcome when prior

knowledge about the shape or movements of the organisms is incorporated into the system.

3.1.1 What is a tracking software?

Among tracking software there is a common workflow in the algorithmic processes implemented, that include from image acquisition to data analysis, and a similar sequence of algorithms (Panadeiro et al., 2021). The identified steps are roughly described below.

Image acquisition: The image acquisition process obtains the images to be processed. The most common approach is offline processing, which extracts a sequence of images from a video file, which are then cleaned and led to the detection algorithms. The primary advantage of offline processing is that, since is not occurring in real time, allows the algorithms to carry out more complex computations with access to future and past frames while processing the current image. Image acquisition has several limitations to be considered. First, it requires special equipment such as experimental setting, lighting and cameras (for specific video quality). Second, these decisions impact the computational cost of the algorithmic processes (speed and memory) needed to run the tracking algorithms. This allows us to see how parameters such as image resolution, codec configuration or video frame rate can affect the computational cost of tracking algorithms exponentially.

Furthermore, for the correct functioning of tracking algorithms, that are based on libraries such as *OpenCV* (García et al., 2015; Gollapudi, 2019), it is recommended that each animal in a video is composed of at least 50 pixels. The recommended frame rate for the algorithms to be able to generate the trajectories depends on the tracking software and the species being tracked. But it should be high enough that the animal position overlaps in two consecutive frames. On the spatial resolution, most tracking applications are limited by the maximum resolution that their detection and tracking algorithms can handle.

Calibration: The calibration process defines the equivalence between the camera's units (pixels) and the real-world units (cm, for example). Some software include algorithms designed to avoid distortions caused by the optics of the lens, misalignment, etc. This step is often unncecesary.

Detection step: This module is where the spatial positions of the animals at each frame are identified. Thus, for every frame in the sequence, the (x,y) coordinates for each detected individual are extracted. Detection algorithms normally include processes such as background subtraction (if all the animals move, stationary features are not a focal point) and segmentation, which reinforces the edges of potential objects. These processes are normally performed one after the other. Some software include a filter in this step to eliminate detected objects that do not satisfy specific conditions, such as shape or size. That is, it removes possible false positives.

The most common segmentation technique uses a threshold value to separate the pixels below and above this determined value. The resulting image satisfies the conditions required by the detection algorithms that require the animals to appear as bright objects in a dark background, or vice versa. This also speeds computational times as the amount of information handled during the detection step is lower.

Trajectory generation: This step takes all the frames where objects/individuals have been detected and creates the movement paths for each animal across sequence of the ordered frames in time. This process addresses the possible occlusions and complex backgrounds or uneven illumination conditions in order to preserve the identity of each individual across its trajectory, while simultaneously avoiding individual swapping. Trajectory generation is a complex task that usually requires manual correction, especially with a large number of animals. Moreover, some software use algorithms to further smooth the trajectories during a post-processing step.

Data Analysis: After the trajectories are created, the data analysis step can be performed. It processes the information from each trajectory to build the desired behavioral data. For example, it is possible to extract values and statistics from moving velocity, mean distance, animal polarization, rate of explored territory or activity rate. There are few software that include this step in their pipeline.

Tracking software limitations

Multiple free tracking software have been developed in the last decade, and understanding which one, if any, suits your experimental settings can be a challenging endeavor. Most tracking applications perform automatic tracking through different approaches and were designed for specific setups, therefore each one has their own strengths and limitations.

It is especially important to check each software requirement in video quality (spatial and temporal resolution) and computational processing power, the species type the application has been tested on, the number of animals capable of tracking simultaneously, individuals' ID preservation through time, and if/how it solves occlusions. For example, some software rely on simple following algorithms that only keep individuals ID's for as long as no occlusions occur (see for example Mönck et al., 2018; Werkhoven et al., 2019). Others, as mentioned in Section 3.1, solve occlusions by analyzing pixels intensity histograms in each individual and joining trajectory fragments by intensity similarity (Rodriguez et al., 2018), and the most sophisticated ones rely on algorithms that search for "visual marks" through deep learning processes to identify individuals or image specific traits (Xu and Cheng, 2017; Romero-Ferrero et al., 2019). The utility of each software depends on the information that needs to be extracted. Experiments that look for only individual frame information or group behavioral features, like area coverage, may not need to invest time in complex algorithms and/or manual intervention for individual tracking, while experiments that require information about individuals velocity or interactions would prioritize ID maintenance and occlusion solving. Table 3.1 summarizes main features of multiple free open-source tracking software (Harmer and Thomas, 2019; Rodriguez et al., 2018; Branson et al., 2009; Rao et al., 2019; Sridhar et al., 2019; Gallois and Candelier, 2021; Walter and Couzin, 2021).

As described in Chapter 2, our black neon tetra videos were recorded with group sizes of 20 and 40 individuals, at medium video quality: a spatial resolution of 1072 x 1004 pixels, each of 8 bits, and a temporal resolution of 20 frames per second. Of the softwares documented, pathtrackr (Harmer and Thomas, 2019) and Animapp (Rao et al., 2019) can only track one individual at a time, while ToxTrac (Rodriguez

Table 3.1: Main tracking software technical features and requirements.

Software	Platform	Computational requirements	Minimum video resolution (px, fps)	Max ind.	Id preservation	Manual intervention
ToxTrac	Windows	8GB of RAM and 2.0+ Ghz Quadcore or higher	1920x1080, 25	20	Yes	Yes
pathtrackr	Windows, Mac, Linux	NA	NA, 30	1	-	Yes
Animap	Windows, Mac, Linux	NA	640x480, 15	1	-	No
idtracker.ai	Windows, Mac, Linux	Core(TM) i7-7700K CPU @4.20GHz 6 core Intel(R) - Nvidia TITAN X / GeForce GTX 1080 Ti / RAM 16Gb-128Gb	1920x1080, 25-50	100	Yes	Yes
Ctrax	Windows, Mac, Linux	32 and 64-bit systems	1920x1080, 20	2-150	No	Yes
Tracktor	Windows, Mac, Linux	Intel i5 2.4 Ghz 8GB RAM	NA, 25	2-8	Yes	No
Fastrack	Windows, Mac, Linux	NA	Any, Any	100	Yes	Yes
TRex	Windows, Mac, Linux	Intel Core i9.790OX CPU / NVIDIA Geforce 1080 Ti / 64 GB RAM	Any, Any	100	Yes	Yes

et al., 2018) and Tracktor (Sridhar et al., 2019) can manage up to 20 individuals. Only Ctrax (Branson et al., 2009) and idtracker.ai (Romero-Ferrero et al., 2019) have the capacity to process over 100 individuals. As we wanted to mantain the identity of the individuals through time, Ctrax was not a viable option, as it does not include ID preservation and was primarily tested in fruit flies. idtracker.ai requires a minimum spatial resolution of 1920x1080 pixels, videos of 25 to 50 frames per second, and high computational power, including a dedicated NVIDIA GPU. We tested our videos on this software, in a computer with Intel Core i7-8550U CPU, 16GB of RAM, an SSD disk and NVIDIA 1650 super GPU. The software was unable to perform the pre-processing step of foreground objects (blobs) detection and orientation. Our videos' spatial resolution was also not enough for AI-based shape identification and occlusions could not be resolved.

New animal tracking software explores new possibilities that are not well covered by existing tools. We found that our videos needed both automatic tracking and an efficient Graphical User Interface to manually correct occlusions and ID mismatches. Therefore, to track the individuals on the videos available for this research, we developed a tracking software, described in Section 3.2. Fastracker (Gallois and Candelier, 2021) and TRex (Walter and Couzin, 2021) are the only software that based

on the description of their technical features, could comply with our requirements. Nonetheless, these software were released in 2021. At this time, the tracking of our experimental data had already been completed.

3.2 Tracking Software Design

To obtain Black neon tetras individual trajectories, a custom made tracking software was built using python *opencv* library (García et al., 2015; Gollapudi, 2019), which makes use of computer vision to detect objects. It can be downloaded from `https://github.com/TheFishTankLab/Peixos_Tracking`, where technical requirements and usage are specified.

In summary, the tracking software consists on the implementation of different processes and algorithms performing the following tasks:

1. Split the video into frames.
2. Detect individuals at each frame using image processing steps such as background subtraction, binarization and filtering.
3. Create trajectories, i.e connecting the detected objects across frames. This is performed by means of a rule system controlling possible candidates in the next frame for each object and a simple assignation algorithm.
4. Run post-processing algorithms to smooth the trajectories.

Assigning each individual at time t to its corresponding position at $t+1$ is monitored by a Graphical User Interface (GUI), described in Section 3.2.6, where a human can inspect the process and manually solve occlusions that can not be solved by the software.

The tracking software requires the user to manually label all the individuals in the first three frames of the video (named $t-2, t-1$ and t). They are used to compute the estimated position of each individual at $t+1$. Then, a rule system is used to select among all the candidate objects at frame $t+1$, those who could actually be the next point in each individual trajectory.

Finally, the algorithm considers all the individuals with their respective candidate positions and iteratively assigns the best candidate until there are no candidates left.

The rule system that selects the candidates aims to capture the expected individual changes in position. It has three rules that limit the distance and direction in which an individual can travel, which are configured by the user. The first two rules limit the possible forward and backward travel distance and angle. The third rule considers a 360° angle within a small radius in the surroundings of an individual i at time t, representing the space where it could be found at $t + 1$. The candidates for an individual are those that satisfy the rules under the parameters selected by the user. The assignation step starts by assigning all the individuals that have only one candidate. Then, the candidates for the rest of individuals are sorted, placing their best candidate first. This candidate is assigned and deleted as candidate for the remaining individuals. This process iterates until no candidates are left and all individuals are assigned to a new position.

We describe how the different steps were solved following the common *taxonomy* for tracking software (Panadeiro et al., 2021)

3.2.1 Image Acquisition

The image acquisition process is described in detail in Chapter 2. In short, we recorded black neon tetra, *Hyphessobrycon herbertaxelrodi* collective motion of groups of 20 and 40 individuals swimming freely in a 100 × 93 × 40cm rectangular experimental tank. The video characteristics are: 20 frames per second, 8 bits of gray levels and a resolution of 1072 × 1004 pixels per frame. This was the maximum resolution the camera was able to provide. The videos consider experimental conditions with a depth of the water column of 5cm (experiment code Q5N20 and Q5N40). Three independent replicates of 12000 frames (coded A,B and C) were recorded for each experimental condition, along with 250 frames of the empty tank for background subtraction.

3.2.2 Detection

In the **detection step** the tracking software finds the (x, y) positions of each fish at each frame. The images are acquired from the videos and processed after by using the *OpenCV* library (García et al., 2015) for computer vision applications.

The detection step includes: frame separation, resizing the background frame if necessary, background subtraction, converting RGB to gray color space and filtering.

- The frame separation process converts the video into individual images.
- The resizing step refers to changing the dimensions of an image, be the width, the height, or both by changing its pixel information. Resizing an image needs a way to calculate pixel values for the new image from the original one. This is usually done with interpolation methods. An interpolation function examines neighborhoods of pixels and optically increases or decreases the size of an image without distorting it. In our tracking software, the new image preserves the aspect ratio of the original image. It uses the *cv2.resize()* OpenCV function using AREA_INTERPOLATION, that resamples using the pixel area relation. For this, first an area of pixels is defined according to the desired modificaton in image dimensions. For example, if we want the new image to be three times smaller than the original in both width and height, then the pixel area will be $3x3 = 9$ pixels. Then, the values of the pixels within it are summed up and then divided by the area. The result is the pixel value of the output image (Ansari, 2020).
- Background subtraction allows to remove stationary features from the environment. This is everything that might be in the background frame, so it is not tracked as a "non-moving" object by the software. The software subtracts the background from the foreground by using a running average technique, that averages pixel values between frames. For this, a set of the video frames are analyzed. In each new frame, the running average is computed over the current frame and the result from the previous frames. Finally, the absolute difference between the background model (which is a function of time) and the current frame is calculated. We implemented this technique via the opencv function

cv.accumulateWeighted(), which includes a parameter to control the weight of the current image frame on the averaging with the accumulated image (Sobral and Vacavant, 2014).

- The RGB to gray color space conversion, together with the filtering, allow to binarize the image, making the objects easier to track. The filtering process includes a Bilateral Filter, and Thresholding. The filter aims to reduce noise by smoothing images while preserving edges. For this, it uses a nonlinear combination of nearby pixel values to define the intensity of each focal pixel. The thresholding takes a grayscale source image, and applies a particular color to every pixel exceeding a threshold value and another color to those below.

The **color space conversion, the filtering and the thresholding** constitute the segmentation step in the common workflow (Panadeiro et al., 2021).

The object detection is performed by using the *Finding contours OpenCV* function, which defines the lines containing all the points alongside a boundary with the same intensity in terms of pixels. That is, the function finds the points where there is a significant difference in the intensity to create contours around. This function works better for binarized images, that is why it is applied after the binarization and filtering processes.

Once all the candidate objects have been detected in a frame, the software generates the trajectories using this information.

3.2.3 Trajectory generation

For the trajectory generation the software connects the detected contours across frames, assigning to each object in a frame the more likely object in the subsequent one to complete the trajectory.

For this propose, the tracking software estimates the position of the fish at frame $t+1$ using the information of the position at $t-2, t-1$ and t. Then, it uses some heuristics, expressed in a set of rules, together with an assignation algorithm, to decide the final position at $t+1$.

The general trajectory generation process can be described as follows:

1. Consider the position of each animal at frame t and detect all possible animals at frame $t+1$.

2. For each animal, estimate its position at frame $t+1$, considering its trajectory ($t-2$, $t-1$ and t).

3. For each animal, define its candidates at $t+1$ as those objects that under certain conditions, defined by the rule system described next, are likely to properly continue the trajectory of the animal.

4. Assign an object at frame $t+1$ to each animal at frame t. This process is carried out by the assignation algorithm, taking into account heuristics such as: assigning first those objects that are the only candidate for an animal, considering proximity and looking at the past trajectory.

Note that the tracking software requires manual identification of individuals for the first three frames, $t-2$, $t-1$ and t, in order to build the trajectories. Each individual identity has to be preserved across these first frames. This process is carried out by means of the user interface provided by the software (described in Section 3.2.6).

Estimating the position at $t+1$

To estimate the position of a fish at $t+1$ the velocity vectors from $t-2$ to $t-1$ as well as from $t-1$ to t are used.

Figure 3.2 shows a schematic of the estimate position process. We proceed in the following way:

1. Calculate $V_{t-2,t-1}$, that is the vector of movement among frames $t-2$ and $t-1$.

2. Place $V_{t-2,t-1}$ over the detected animal at frame $t-1$. The result is the estimated position of the animal A at frame t. This considers that the animal moves along the same direction and with the same speed.

3. We estimate the error E at t considering the real position of the animal at frame t and the estimated position. The error corresponds to a change of direction

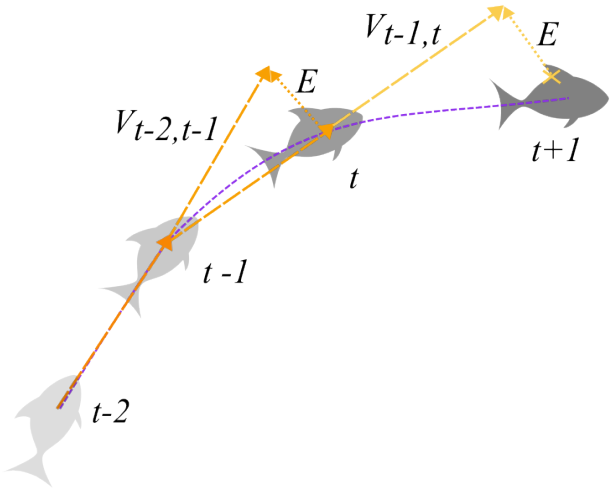


Figure 3.2: Schematic of the process to estimate the position of a fish at frame $t+1$, based on its positions at frames $t-2$, $t-1$ and t. The process builds the vectors of movement from $t-2$ to $t-1$ and from $t-1$ to t, named $V_{t-2,t-1}$ and $V_{t-1,t}$.

(i.e if the animal is turning to the left or to the right) or speed, and express the variations in comparison to the previous frame.

4. Calculate $V_{t-1,t}$, that is the movement of the animal between $t-1$ and t.

5. Place $V_{t-1,t}$ into the center of the position of the animal at frame t. Then add E to the result. The resulting position corresponds to the estimated position of the animal at frame $t+1$.

This estimation of the position intends to capture the curved paths that the fish tend to follow, considering the direction of the velocity into the prediction of the next position.

Calculating the estimated position helps to solve problems such as the following one: when an occlusion occurs, the closest detected object at $t+1$ will be a candidate for all the individuals occluding. However, if we estimate their positions at $t+1$ considering its trajectories it is likely that we have two or more (depending on the number of individuals) different estimated positions at $t+1$ from which to proceed.

Rule system

Now that we have an estimated position for each fish at $t+1$, we analyze which of the detected individuals at $t+1$ are actually good candidates for each estimated fish position at $t+1$. All the fish with their selected candidates will be then processed by

an assignation algorithm. Selecting candidates allows faster processing as not every individual at frame $t + 1$ has to be considered in the assignation process.

The rules are based on heuristics intended to capture the physical movement of the animals and to handle small variations that are produced by the *OpenCV* detection function.

The rules can be tuned by the user as well as activated or not according to the performance of the tracking software for a specific video. For example, there is a rule to capture the fact that the species being tracked *does not move backwards*. This rule allows to discard objects located within certain angle which, according to the biology of the species, is unlikely to be reached by an individual from one frame to the next. Rule parameters are defined by the user in a configuration file. Next, we describe the configurations and the rules.

Rules Configuration

Rules can be configured at a *config.ini* file. The file requires the following values:

- dis_rule_front: an offset to adjust the tracking zone.
- max_dist_front: the maximum distance that an animal can move forward in one frame. Objects beyond this distance are not consider as valid candidates.
- front_angle: the angle within which it is consider that an animal can move forward.
- dis_rule_behind: an offset to define the tracking region.
- max_dist_behind: the maximum distance that an animal can move backwards in one frame. Objects beyond this distance are not consider as a valid candidates.
- dis_rule_inner: offset to configure the tracking region.
- max_dis_rule_inner: maximum distance that we consider an animal can move in any direction. Objects beyond this distance are not consider as a valid candidates.

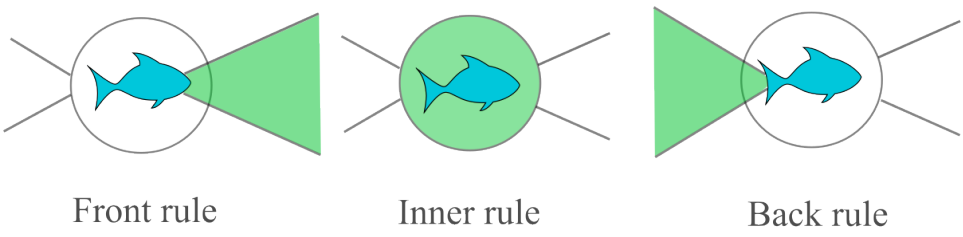


Figure 3.3: Angles of the front, inner and behind rules define a 2 dimensional space where the candidate objects can be.

The offsets dis_rule_front, dis_rule_behind and dis_rule_inner, allow to define a region beyond the closest object detected to search for candidates. The maximum distances max_dist_front, max_dist_behind and max_dis_rule_inner bound the search in case the closest detected object is too far. This means that if the sum of the closest detected object and the offset is greater than the maximum distance, then a fish has no candidates for that rule.

There is a global parameter *max_long_jump* set in the configuration file, which monitors "long jumps". This is, displacements considered impossible for the species being tracked, that could result of an unusual speed increase of an individual, or an error in the individual assignation. This parameter is intended to speed the process of manual correction, and to simplify the tracking process.

These rules parameters are used for three rules. Named: Front, Inner and Behind rule. Each of them has an angle and threshold distances for being able to select different areas. A visual description of the angles of the rules: Front, Inner and Behind, can be found at Figure 3.3.

Front Rule

This rule aims to capture fish forward movement. It is important to remark that we are measuring position changes in $\frac{1}{framerate}$ seconds, so for standard video temporal resolutions, the change in position among two consecutive frames is small.

Let $P_{A,t}$ be the position of fish A at time t, and $P_{A,t+1}$ the estimated position of fish A at $t+1$. Let min_dist be the distance between the closer detected object at frame $t+1$ and $P_{A,t+1}$. Then, an object B at frame $t+1$ is a candidate for A according to the front rule if it satisfies the following conditions:

$$dist(P_{A,t+1}, B) < \text{min(min\_dist + dis\_rule\_front), max\_dist\_front)} \text{ and angle} < \text{front\_angle} \quad (3.1)$$

where angle corresponds to the absolute value of the angle between the vector from $P_{A,t}$ to $P_{A,t+1}$ and the vector from $P_{A,t}$ to B, and front_angle corresponds to the angle allowing forward movement.

The value dist_rule_front defines how much, beyond the distance of the closer object detected at $t+1$ (min_dist), the software will look for possible candidates B. If we only consider the closest distance we would not consider other plausible candidates. Such distance is bounded by the max_dist_front.

Behind Rule

This rule is designed to limit the backward movement. Also, it is intended to improve the tracking process when an animal remains static but the object detection library places the detected object "backwards", as it identifies the centroid of the detected blob, but the size and/or shape of the blobs can vary from frame to frame, due for example to changes in fish tilting or startled fish that appear longer.

Let $P_{A,t}$ be the position of fish A at time t, and $P_{A,t+1}$ the estimated position of fish A at $t+1$. Let min_dist be the distance between the closer detected object at frame $t+1$ and $P_{A,t+1}$. Then, an object B at frame $t+1$ is a candidate for A according to the behind rule if it satisfies the following conditions:

$$dist(P_{A,t+1}, B) < \text{min((min\_dist + dis\_rule\_behind), max\_dist\_behind)} \text{ and angle} < \text{behind\_angle} \quad (3.2)$$

where angle corresponds the absolute value of the angle between the vector between $P_{A,t}$ and $P_{A,t+1}$ and the vector going from $P_{A,t}$ to B, and behind_angle corresponds to the angle where the animal can move backward.

As in the case of the front_rule, the value dist_rule_behind defines the distance beyond the minimum distance in which the objects that are found will be included as candidates. max_dist_behind limits the maximum distance where a detected object can be considered as candidate for A.

Inner Rule This rule considers movements that can be produced in any direction. This rule also helps to consider small variations during the tracking process due

changes in the position of the centroid of the detected blob from frame to frame.

If $P_{A,t+1}$ is the estimated position of fish A at $t+1$ candidates according to the Inner Rule will be blobs B satisfying:

$$\text{dist}(P_{A,t+1},\text{B}) < \min((\text{min\_dist} + \text{dis\_rule\_inner}), \text{max\_dist\_inner}) \tag{3.3}$$

Again dis_rule_inner defines how far from min_dist will the objects found included as candidates.

Each rule creates a set of possible candidates. The candidates for a fish at $t+1$ will be the union of the sets containing the candidates selected by each rule.

The process of selecting the candidates for a fish at $t+1$ is described pseudocode in Algorithm 1.

Algorithm 1 Get candidates function

function GET CANDIDATES(positions at t-2, t-1, t, t+1, dis_rule_front, max_dis_rule_front, front_angle, dis_rule_behind, max_dis_rule_behind, behind_angle, dis_rule_inner, max_dis_rule_inner)
    for each animal *a* in frame t **do**
        Compute the estimated position at $t+1$ considering $t-2$, $t-1$, and t. This process is described in detail in Section 3.2.3.
        Compute the distances from the estimated position of the individual to the objects detected at $t+1$
        Let "distance" be the distance from the estimated position (at $t+1$) to a current candidate.
        Let min_dist be the minimum of all the distances.
        Candidates for *a* are objects satisfying:
        front_rule_active and distance $<$ min((min_dis + dis_rule_front),max_dis_rule_front) and angle $<$ front_angle
        behind_rule_active and distance $<$ min((min_dis + dis_rule_behind), max_dis_rule_behind) and angle $>$ behind_angle
        inner_rule_active and distance $<$ min((min_dis + dis_rule_inner), max_dis_rule_inner)
    end for
    return Candidates
end function

3.2.4 Assignation Process

The function that collects the candidates information and assigns to each detected object at time t its best candidate at $t+1$ operates as follows. For a fish, the "best

assignation" is considered to be its closest object, given that the system has already considered the direction of movement for three frames to estimate its position at frame $t + 1$. Moreover, the rules have already filtered objects that are unlikely to be the next position on its trajectory. In this way, the *best assignation function* finds for each fish its closest object considering these restrictions. The function starts by assigning those fish that have only one candidate. Once this is done, it assigns the rest of the fish to its closest object among the candidates. And finally, it assigns those individuals that might have no candidates (the rules could return empty sets) to its closest object considering the raw data as it is captured before being processed by the rules.

3.2.5 Post-processing of tracking data

Opencv detects blobs and assigns the position of the detected object to the center of its darkest region. However, when a fish remains steady for a few frames, blob variations can be produced in the subsequent frames, creating, given detection differences, an unnatural motion where the position moves slightly forward and backwards.

The Savitzky-Golay filter is a commonly used filter in signal processing to smooth data, i.e. to increase the signal precision without distorting its tendency (Savitzky and Golay, 1964). It calculates a polynomial fit of order N by linear least squares in a sequence of points.

Roughly speaking, the Savitzky-Golay filter proceeds in the following way for a given $x[n]$ signal of successive points. Consider a window of size $2M + 1$ points centered at point n^*. It is possible to obtain the coefficients of the polynomial:

$$p(n) = \sum_{k=0}^{N} a_k n^k \tag{3.4}$$

that minimize the least squares error of the window. i.e.

$$\epsilon_N = \sum_{n^*-M}^{n^*+M} (p(n) - x[n])^2 \tag{3.5}$$

The value of the smoothed function would be $p(n)$ evaluated at n^*. That is, $y[n^*]$, the output at $n = n^*$ is

$$y[n^*] = p(n^*) = a_0 \tag{3.6}$$

i.e., the output value is equal to the 0th polynomial coefficient. The filter computes the output value of the central point by shifting the window by one point at a time during the length of the signal.

3.2.6 Graphical User Interface

The tracking software provides a Graphical User Interface (GUI) where the user controls the tracking process. When the GUI is initiated, the software first asks the user to set the number of individuals, the path of the background video (otherwise it is calculated automatically), and the path to a specific tracking file (otherwise it initiates a new file). After those variables are set, the GUI changes to the one shown at Figure 3.4 where the tracking process is supervised. At the left bottom of Figure 3.4 there is a bar that allows frame to frame shifting. In the example image, we are positioned at frame 391. The same frame to frame movement is allowed by the small box located an the end of the bar, also holding number 391. To assign an ID to a particular fish, the **New Position From** box is used. It allows to select a number among the individuals being tracked and assign it to the fish selected by the mouse pointer when the left click is pressed.

When assigning the id to the first thee frames, the bottoms at **Display Options to facilitate Manual Intervention** region can be used. By selecting the **Show Only One Fish** option, the tracking software automatically assigns the Selected id across frames. If the software switches ID's after two fish occluding, the **Switch from** button allows to swap selected identities.

Buttons **New Position** changes the ID position of the selected fish at the current frame without altering subsequent frames, and **Full New Position** changes the fish ID position in a particular frame and resets that ID positions information in all the subsequent frames. This second option is used when we need to correct a full trajectory where an error is being propagated across frames.

Button **Save** writes the x,y positions by frame into a python pickle file, and button **Post Processing** applies the smoothing algorithm to the saved data. Finally, the button **Play Automatic assignation** calls the tracking algorithm to automatically assign frame by frame the positions of the individuals, and it will stop if the box **Stop if Not Assigned** is checked to allow the user to manually correct the unassigned fish. The error area is circled if the **Show Error Zone** box is checked.

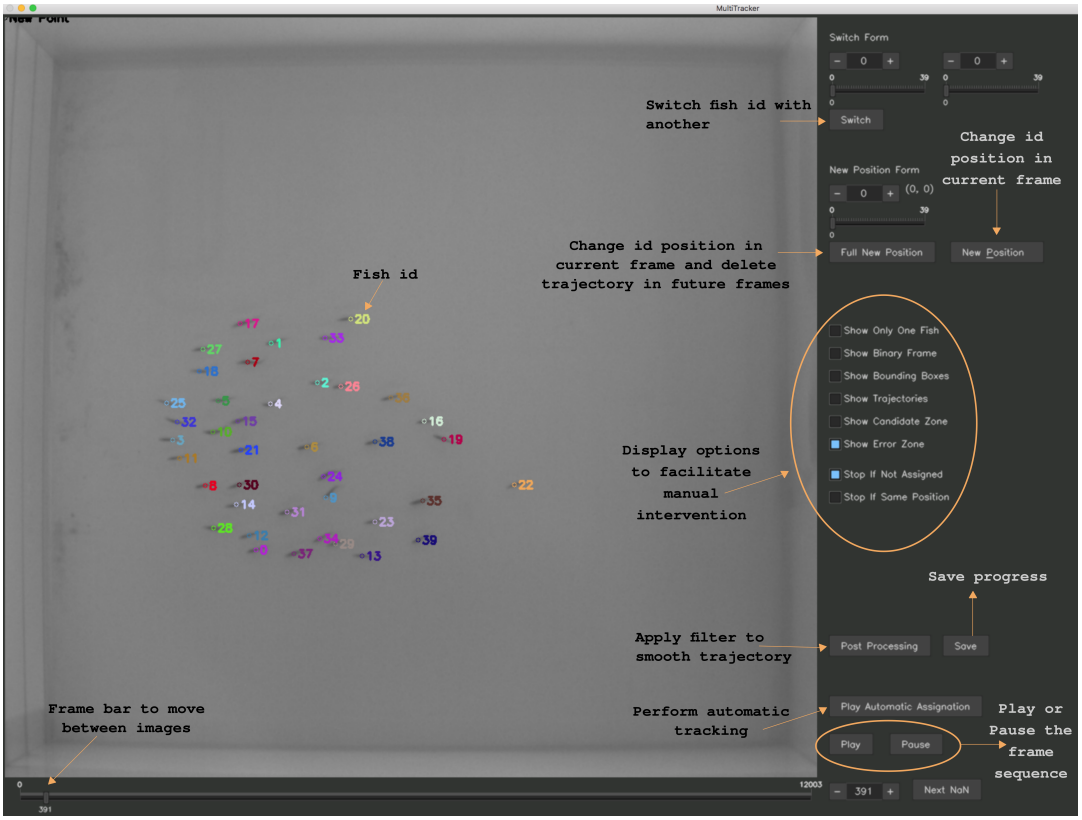


Figure 3.4: A: Image of tracking Software Graphical User Interface (GUI) with main controls for automatic and manual tracking highlighted.

3.3 Tracking results

In total, three 12,000 frames replicates ($\approx$ 10 min videos) per experimental condition (Q5N20, Q5N40) were processed, adding up to 72,000 frames.

For each video, first background subtraction was performed by analyzing all 12,000 frames and removing stationary features corresponding to the tank edges and black spots on the tank floor. Figure 3.5, A shows an example of the background obtained by this process. Next, parameters were tuned to achieve a correct image

segmentation by thresholding, filtering and binarization. For this, we used a threshold value of 30 (of gray-scale from 0 to 255), which resulted in clear fish shapes where pixels where darker than the threshold. We can check the resulting binarization directly in the GUI, by enabling the **Show Binary Frame** box (Figure 3.5, B).

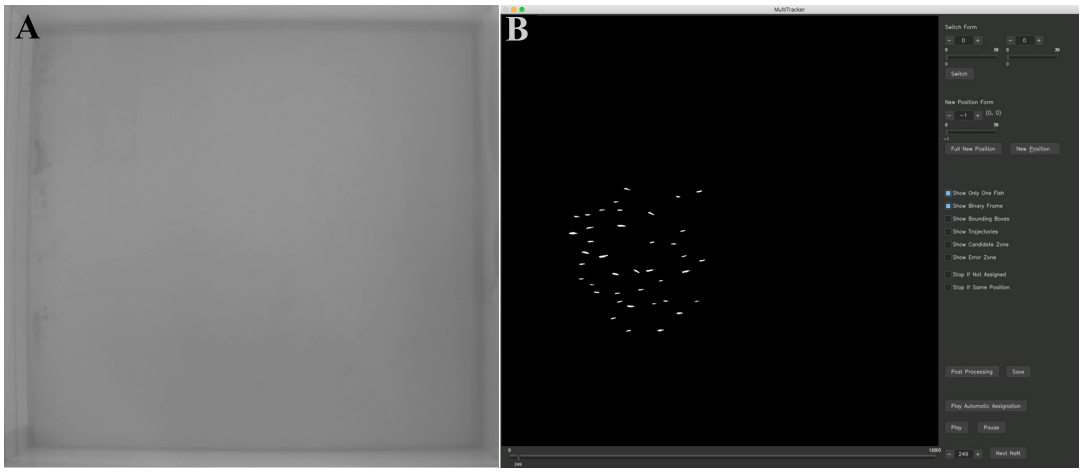


Figure 3.5: **A**: Background subtraction resulting image. **B**: Software GUI visualization of binarization process.

For the trajectory generation, we also tuned parameters by direct observation of object detection of approximately 100 frames with different parameter combinations. Finally we set our parameters for the rule system as shown in Table 3.2, and proceeded with the tracking by automatic assignation of a bounding box to each detected shape. We kept the **Show Error Zone** and **Stop if Not Assigned** boxes checked so the software would show us whenever a fish had no good candidates according to the rule system parameters, either caused by rare long jumps or because it was impossible for the software to differentiate between two or more fish after an occlusion (See example in Fig. 3.6,B). These circumstances where solved by manual

Table 3.2: Parameters used for the tracking of 20 and 40 individuals of black neon tetra.

Rule	Value	Unit
dis_rule_inner	3	px
max_dis_rule_inner	15	px
dis_rule_front	6	px
max_dis_rule_front	35	px
dis_rule_behind	4	px
max_dis_rule_behind	15	px
front_angle	15	degrees
behind_angle	170	degrees

intervention. Complete tracking of each video took from 6 to 10 hours, depending on the need for manual intervention, divided in 2-hours work sessions. Example GUI images of the trajectories generation process are included in Figure 3.6,C).

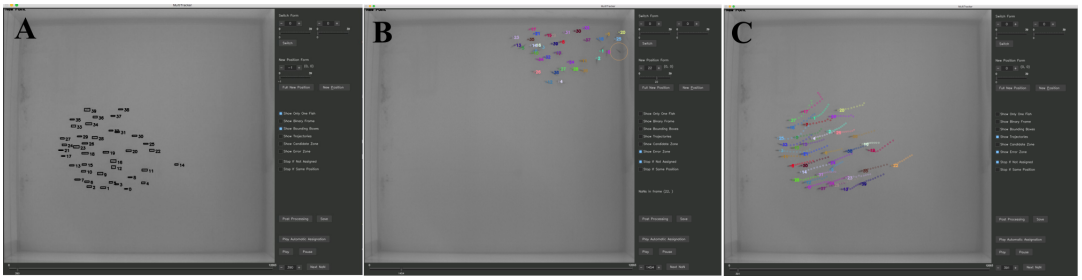


Figure 3.6: **A**: Bounding boxes assignation in detection step. **B**: **C**:

Once all the trajectories were completed, a a post-processing step was included to correct small path deviations due to either manual intervention or different shape size detection, and hence different location of the centroid. For this, we applied a Savitzky-Golay filter of degree three over a 0.35 s time window (7 frames; Fig. 3.7). Finally, we obtained the trajectories in a .csv file with frame, fish ID, and x, y coordinates in pixels. We then transformed x,y coordinates from px to cm according to image resolution ($1072x1004px$) and tank measures ($100x94cm$).

3.3.1 Software performance

To explore the obtained trajectories we first plot the normalized density of fish space occupation in the experimental tank. In Fig 3.8 we show replicas A,B and C of condition Q5N40 as example. Fish moved with more frequency in central tank areas and less near the tank walls.

We calculate individual and collective behavior descriptors to validate our tracking results, and compare measures with other studies that have worked with black neon tetras or closely related species, using different tracking software and methodologies.

First, to check the coherence in each individual trajectory, we calculated the individual instantaneous speed,

$$v_i(t) = |\vec{v}_i(t)| \equiv [\vec{v}_i(t) \cdot \vec{v}_i(t)]^{1/2}, \tag{3.7}$$

and turning angle per frame, defined as the angle of rotation from $v(t)$ to $v(t+1)$,

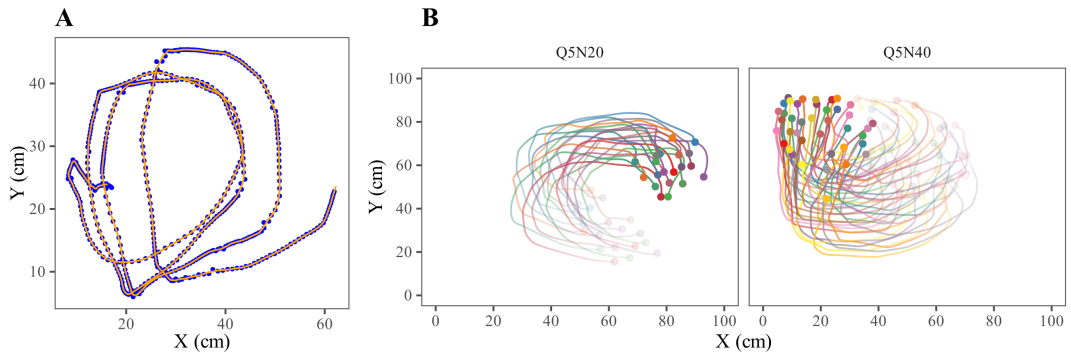


Figure 3.7: **A**: Example of pre-processed fish trajectory (blue points) and smoothed trajectory after applying Savitzky-Golay filter (orange line). **B**: Final trajectories of 1000 frames tracked fish in replica B of conditions Q5N20 and Q5N40.

$$\varphi_i(t) = \left\| \arctan \left\{ \frac{\|\vec{v}_i(t) \times \vec{v}_i(t+1)\|}{\|\vec{v}_i(t) \cdot \vec{v}_i(t+1)\|} \right\} \right\|, \tag{3.8}$$

where $\times$ stand for the vectorial product and $\|\cdot\|$ represents the vector modulus. The angles are computed in the interval $[-\pi, \pi]$ to check for symmetry.

Table 3.3 shows mean speed values and standard deviations aggregated for all individuals, per each tracked experiment. Individual speed values are in tune with reported speeds in other fish studies when divided by body length (Palstra et al., 2010; Videler and Wardle, 1991), and with closely related species *H. rhodostomus*, that move with a mean speed of 0.7 cm/frame when swimming freely in an experimental tank (Calovi et al., 2018).

Additionally, each trajectory's speed in time exhibits a burst-and-coast swimming pattern, characterized by an active phase during which the fish accelerated (burst phase), followed by a passive phase where they glide and keep their bodies

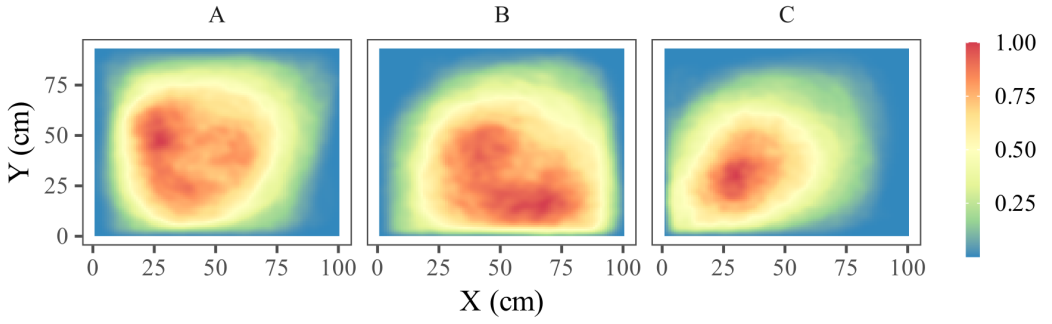


Figure 3.8: **A**: Normalized density of space occupation of 40 black neon tetra swimming freely in an experimental tank. Replicas A,B and C of condition QN540 are shown.

Table 3.3: Mean individual speed values, $\langle v_i \rangle$, and mean individual distance to the nearest neighbor, NND, for each replicate and condition tracked.

Condition	**replicate**	$\langle v_i \rangle \pm S.D.$ $(cm/frame)$	$\langle NND \rangle \pm S.D.$ (cm)
Q5N20			
	A	0.89 ± 0.42	4.01 ± 1.48
	B	0.83 ± 0.42	4.10 ± 1.50
	C	0.87 ± 0.44	4.30 ± 1.62
Q5N40			
	A	0.92 ± 0.34	3.33 ± 1.29
	B	0.94 ± 0.37	3.78 ± 1.41
	C	0.89 ± 0.38	3.79 ± 1.44

motionless and straight (coast phase) (Videler and Weihs, 1982). This swimming pattern is also observed in other fish species with body and caudal fin propulsion (Fish, 2010b), such as *H. rhodostomus* (Calovi et al., 2018). The burst-and-coast swimming behavior in black neon tetras is explored in more detail in Chapter 5.

Individuals turning angles can be used to analyze trajectories smoothness and video resolution. We would expect that our resolution (20 frames per second) captures fish continuous motion, which would result in mostly smooth directional changes (turning angles close to 0) and occasional larger turns when individuals abruptly change direction. The distributions of $\varphi_i(t)$ for each replicate are plotted in Figure 3.9. Turning angles distribute symmetrically between $[-\pi, \pi]$, and the high frequency of $\langle \varphi_i \rangle$ values close to zero falls within expectations of smooth trajectories with small directional changes. Larger turns are also represented with low frequencies. In fact, 95% of observed turning angles fall between ± 0.429 radians.

One of the most commonly used parameters to measure shoal behavior, i.e. to assess group cohesion, is the distance of the focal fish to the nearest neighbor (NND). Here, we observed mean NND values from 3.33 to 6.14 cm (Table 3.3), that correspond to 1.3 to 2.2 body lengths (black neon tetra's body length is of $\approx$ 2.5 cm, Chapter 2) which is consistent to observed NND in black neon tetra (Quera et al., 2019) and other species with different body length and group sizes (Katz et al., 2011; Herbert-Read et al., 2011). To verify that our trajectories depict the schooling tendency of interacting black neon tetras, and compare results with observations in this species, we calculate the heading difference in time between the focal fish and it's

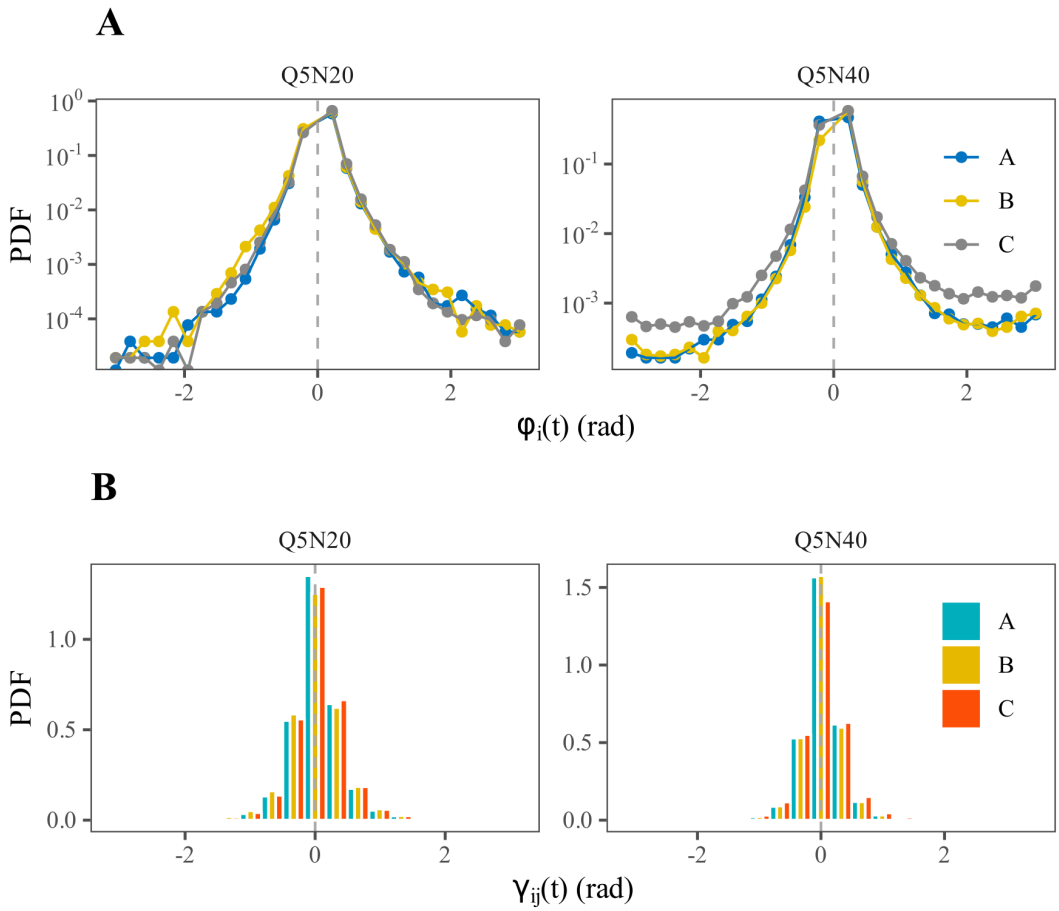


Figure 3.9: A:Probability density function of individuals turning angle $\langle \varphi_i \rangle$ for A,B and C replicates of each condition. **B**: Probability density function of individuals heading difference with their closest neighbor, $\gamma_{ij}(t)$, for A,B and C replicates of each condition.

nearest neighbor,

$$\gamma_{ij}(t) = \left| \arctan \left\{ \frac{\|\vec{v}_i(t) \times \vec{v}_j(t)\|}{\|\vec{v}_i(t) \cdot \vec{v}_j(t)\|} \right\} \right|, \tag{3.9}$$

with angles also computed in the interval $[-\pi, \pi]$.

Figure 3.9, B shows $\gamma_{ij}(t)$ distributions for each replicate. As with turning angles, we observe angle symmetry in all experiments, and a high frequency of small heading differences, which indicates that neighbouring fish heading tends to be similar to that of the focal fish, i.e. a high polarization and individuals schooling tendency, also reported in (Quera et al., 2019).

3.4 Closing remarks

The need for animal behavioral data, specially about the way they move, has elicited numerous tracking software solutions that tackle various behavioral and/or technical specifics. Nonetheless, both the stochasticity of animal systems and the array of experimental designs and/or resources available makes most software non-transferable to use under different experimental settings. We found that only a couple of applications could track more than 20 individuals at a time, and from them, the only software that could preserve identities through time needed a high video resolution and expensive computational power, which were not available for our research. We built a software that allows to automatically track at least 40 individuals following similar image-processing approaches to other tracking software, and a rule-system algorithm that aims at estimating future individual positions based on previous directional behaviors and on knowledge of common maximum distance traveled from frame to frame. As our videos were recorded at medium resolution, precluding individual fingerprinting, we built an intuitive, functional GUI to perform manual corrections when position change identification by the algorithm was not possible, and corrected possible path deviations with a noise filtering algorithm.

We obtained clear individual trajectories for each experimental condition recorded, and categorized fish individual features to compare our tracking results with behavioral traits of black neon tetras and closely related species reported elsewhere.

4 Modeling of collective motion

4.1 Models of collective motion

Collective motion is a widespread phenomenon that can be observed in systems of very different nature, such as bird flocks, fish schools or human crowds. An intriguing property of collective motion phenomena is that even in very different systems, be cells clusters or human crows, very similar collective behaviors can emerge under certain conditions (Lopez et al., 2012). Following this observation, a collective motion system has been conceived as a collection of individual units that interact in relatively simple ways. This idea faciliates the use of theoretical modeling to disentangle the underlying interactions from which collective motion emerges. In the last decades, multiple models have been proposed to simulate the most characteristic features observed in natural systems exhibiting collective motion. These models have contributed to the idea that there are universal mechanisms by which individuals' local rules of interaction generate the emergent global patterns observed.

In theoretical approximations, the universal laws that could apply to collective motion systems are being investigated with models of self-propelled particles (SPP; Deutsch et al., 2012). Although in recent years the increase in theoretical and experimental studies has deepened our understanding of collective motion, the proximate causes of the collective patterns diversity are still poorly understood (Herbert-Read, 2016; Vicsek and Zafeiris, 2012). The philosophy of SPP models is to provide the simplest possible model that reproduces the key features of flocks, herds, migrating cells, etc. The purpose is to ensure that model outcomes are not dependent on some particular system, and hence to reveal the universal properties of the global behavior exhibited by the individuals. Moreover, SPP models allow us to investigate the conditions under which collective patterns are produced by spatially local interactions. Establishing a connection between reproducing the global behavior of

collectives and describing how between-individual interactions produce group dynamics is where SPP models play an important role.

In SPP models particles move in a one, two or three-dimensional space. Each particle has a local interaction zone within which it responds, i.e. affects and is affected, to other particles. The exact form of this interaction varies between models but, typically, individuals are repulsed by, attracted to and/or align with other individuals within one or more different zones. Typically, noise is also incorporated into the movement of the individuals to replicate the stochasticity of real world systems.

Collective motion can be expressed as an ordered motion in one extreme, and as disordered random motion on the other. In models of SPP, a *transition* from disordered to ordered motion can emerge by tweaking relevant parameters of the model. As discussed earlier in Chapter 1, a *phase transition* is a process from which the collective of many interactive particles undergoes a transition from one *phase* to another depending on certain parameter values (Lopez et al., 2012). Phase transitions are defined by the change of a parameter that indicates the level of order that characterizes a phase, with typically zero indicating a disordered phase and one an ordered phase. The way the order parameter changes provides information about the possible states in which the system can be found. Generally speaking, a discontinuous change in the order parameter is referred to as a first order phase transition, on which both phases can coexist, while a continuous change is called a second order phase transition (Vicsek and Zafeiris, 2012).

The simplest SPP model was proposed by Vicsek et al., 1995. It is defined by a group of N identical particles moving at a constant speed in a d-dimensional space. At each time step, they change their orientation depending on both the average orientation of their neighbors (where are they heading) and a noise term that randomizes the direction of motion of each particle.

Based on this model, others have been built with varying levels of sophistication regarding the interaction rules. Examples of this include changing the number and properties of neighbors that the focal individual interacts with (Camperi et al., 2012), varying the type of interaction depending on the particles distance and direction (Wang et al., 2013), or considering the effect of spatial memory on group distribution (Grünbaum, 2012).

For example, the model proposed by Couzin et al., 2002 is based on a series of local repulsion, alignment and attraction rules according to the particles' relative position and alignment. These rules assume that individuals operate in a hierarchical manner. First, they try to maintain a minimum distance with others at all time. Second, individuals tend to be attracted to other individuals. Finally, they try to align with each other.

In this chapter, we describe the Vicsek model with more detail and explore some of its properties while varying the control parameters that characterize the model. We specially focus on particles' speed-based phase transitions behavior, as is the least explored property in the literature. In Chapters 5 and 6 we work with modified versions of the Vicsek model to explore various emergent collective properties in our experimental fish schools.

4.2 The Vicsek Model

To establish a quantitative interpretation of flocking behavior the Vicsek model was introduced in 1995 (Vicsek et al., 1995). The Vicsek model describes the dynamics of a set of self-propelled particles (SPP) characterized by an instantaneous position $\vec{x}_i(t)$ and direction of motion given by the unit vector $\vec{s}_i(t)$. The dynamics is overdamped and the position of particles is updated in a parallel scheme in discrete time as:

$$\vec{x}_i(t+1) = \vec{x}_i(t) + v_i(t)\vec{s}_i(t), \tag{4.1}$$

where $v_i(t)$ is the speed of particle i at time t. Periodic boundary conditions in the box containing the particles are imposed.

Interactions in the Vicsek consist in the alignment of the velocity of an SPP with the average velocity of some of its neighbors. Perfect alignment is, however, impeded by the addition of a noise term that mimics, for instance, the difficulties in gathering or processing the surrounding information. In two dimensions, assuming a direction of motion in polar coordinates, $\vec{s}_i(t) = (\cos(\theta_i(t)), \sin(\theta_i(t)))$, in terms of

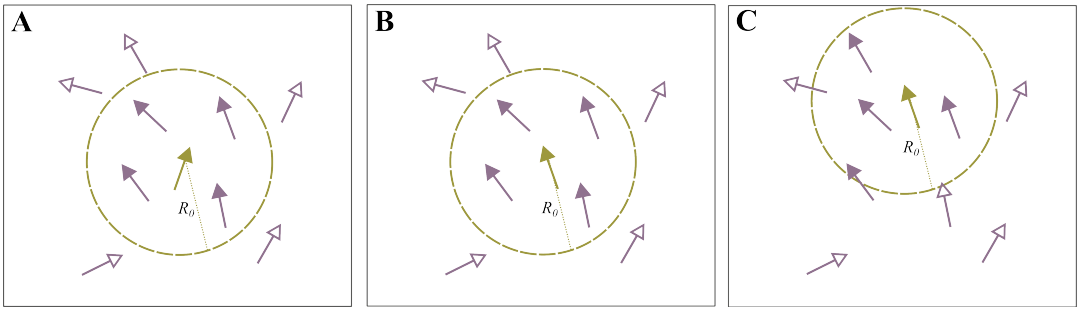


Figure 4.1: Diagram of the Vicsek model dynamics. **A**: The focal particle (green) has three neighbors (colored purple) within a radius R_0. **B**: The focal particle changes its orientation averaging its neighbors' orientations. Noise prevents perfect alignment. **C**: The focal particle moves at a constant speed v_0 towards the new direction of motion. In the Vicsek dynamics all particles change orientation and position synchronously.

the angle θ with the horizontal axis, the orientation dynamics is expressed as (Vicsek et al., 1995)

$$\theta_i(t+1) = \text{Arg}\left[\sum_{j\in\mathcal{N}_i(t)} \vec{s}_j(t)\right] + \eta\xi_i(t), \tag{4.2}$$

where $\text{Arg}\left[\vec{V}\right]$ represents the angle described by vector $\vec{V}$, $\mathcal{N}_i(t)$ represents the set of particles at a distance smaller than or equal to R_0 of particle i (including i itself) at time t, $\xi_i(t)$ is an uncorrelated random noise uniformly distributed in the interval $[-\pi, \pi]$, and $\eta \in [0, 1]$ is a parameter representing the noise strength. The model is thus fully defined in terms of the number of particles N, the size of the simulation box L (related by the density $\rho = N/L^2$), the radius of interaction between particles R_0, the noise strength η, and the particle speed $v_i(t)$. In the classical Vicsek model, speed is constant and the same for all particles, $v_i(t) = v_0$.

A typical choice to reduce the number of variables, is to select $R_0 = 1$. Then, the model only depends on parameters η, v_0 and ρ.

Equation 4.2 contains a explicit alignment term that *competes* with the noise parameter η to define the orientation of each single particle. When it is strong enough to overcome the effect of the noise, the system develops a global order, and the particles appear oriented in the same direction generating collective motion. This is measured by a polar order parameter, i.e. the average normalized velocity:

$$\phi(t) = \left|\frac{1}{N}\sum_{i=1}^{N} \vec{s}_i(t)\right|$$

The polar order stationary time average, $\langle\phi\rangle = lim_{T\to\infty}\frac{1}{T}\int_0^T \phi(t)dT$, is used to describe order transitions, with $\langle\phi\rangle > 0$ representing the ordered phase.

4.3 Properties of the Vicsek Model

One of the main attractives for the statistical physics community to the study of the Vicsek model is the behavior of the phase transition when the system's behavioral parameters are varied (Aldana et al., 2009).

Vicsek et al., 1995 showed that, when η is increased, the system undergoes a phase transition from an ordered state where all the particles move in the same direction ($\langle\phi\rangle > 0$), to a disordered state where the particles move in random directions ($\langle\phi\rangle \simeq 0$).

As for the nature of the phase transition, numerical results first showed that a second order (continuous) transition ocurred at a critical η value (Vicsek et al., 1995). Figure 4.2 shows the stationary order parameter as a function of noise amplitude for different system sizes, where a continous transition is apparently observed. In the thermodynamic limit, the order parameter is expected to change according to

$$\langle\phi\rangle \approx [\eta_c(\rho) - \eta]^\beta, \tag{4.3}$$

where β is a critical exponent, and $\eta_c(\rho)$ is the critical noise for $L \to \infty$ depending on ρ.

These results were later challenged by numerically analyzing large systems with many more particles moving at higher speeds. From these analysis a first order (discontinuos) transition was observed (Grégoire and Chaté, 2004). This elicited a debate about the nature of the transition in the Vicsek model, and it has resulted in a number of studies investigating the order-disorder phase transition under different parameter combinations (Aldana et al., 2009, Baglietto and Albano, 2009a, Clusella and Pastor-Satorras, 2021).

The main findings are that both the critical value at which the phase transition occurs and the nature of the transition are dependent on combinations of speed, density and system size, which act as regulators of the particles' interactions, and hence, of the randomness in the direction of motion. For low densities and low particles

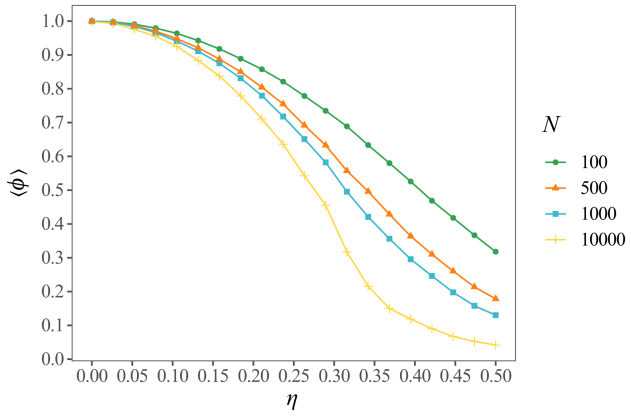


Figure 4.2: $\langle\phi\rangle$ as a function of η for different number of particles, keeping $\rho = 0.5$. $\langle\phi\rangle$ was averaged over 100,000 time steps (5,000 time steps of thermalization) in three simulations repetitions.

speeds, the phase transition is predicted to occur at low noise amplitudes, regardless of the system size, and to be compatible with a second-order phase transition (Sumpter et al., 2008; Aldana et al., 2009), as shown previously in Figure 4.2. Under this regime, the transition is fairly independent of the speed value as shown in Figure 4.3 where different speed values produce a very similar order-disorder transition curve. There is an exception for $v_0 = 0$, where no transition is observed as noise is increased.

When density and particles' speed are high enough and the system size is large, the predicted noise critical value is higher, and the order-to-disorder phase transition is closer to a first-order, discontinuous phase transition (Grégoire and Chaté, 2004; Aldana et al., 2009).

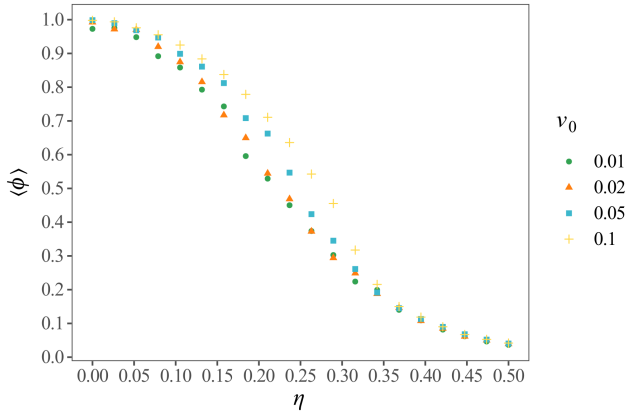


Figure 4.3: Stationary ϕ as a function of η in the Vicsek Model for multiple values of v_0.$\langle\phi\rangle$ was averaged over 100,000 time steps (5,000 time steps of thermalization) in three simulations repetitions.

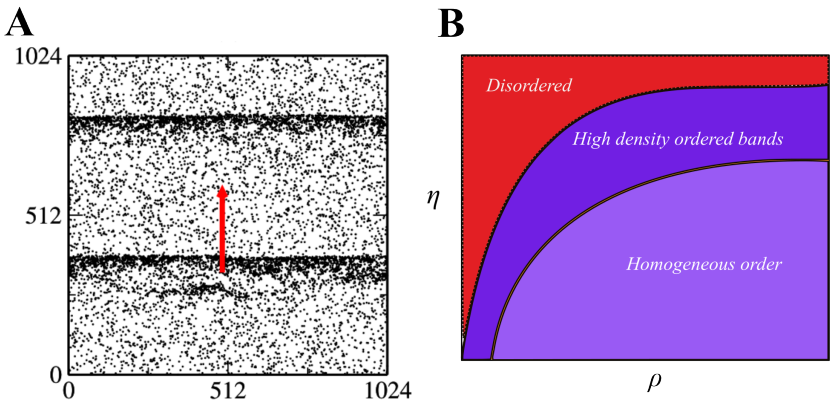


Figure 4.4: **A**: High density particles' bands within a disorganized Vicsek system. Image borrowed from Ginelli, 2016. **B**: Qualitative phase diagram of the type of order in the Vicsek model with combinations of ρ, η.

An interesting property of the Vicsek model is the heterogeneous particles' spatial distribution that can emerge in the system due to the interplay between local order and local density in the first order regime. For a combination of noise amplitudes and densities, particles may gather in high density bands. These patches create an increase in the neighbors that a particle interacts with and as a result, local alignment improves while the rest of the system is desorganized. This mechanism generates a long wavelength instability near the onset of order, that breaks the homogeneous ordered phase and leads to spontaneous phase separation. In Figure 4.4 we can see a snapshot of particles arranged in high-density bands with a well-defined width moving through an otherwise disorganized system. When noise amplitude decreases and particles' density increases, the instability disapears and an ordered phase emerges, as can be seen in Figure 4.4B, that shows a schematic of the three phases of the Vicsek model: A disordered phase, where particles act as persistent random walkers at high noise amplitude, a phase separated ordered system with high-density bands, and an homogeneous ordered phase for high densities and low noise amplitudes (Ginelli, 2016).

Although the model has been extensively analyzed varying the noise amplitude and the particles' density, to our knowledge there is very few literature related to the order-disorder phase transition as a function of the particles' speed. The single study whose objective was to understand the speed-induced ordering phase transition worked with a modified version of the Vicsek model that included an attraction

radius to keep particles' cohesion (Cambui and Rosas, 2018). They observed that increasing particles' speed caused a decrease in the order parameter, and the nature of the order-disorder phase transition depended on the noise intensity and the attraction radius between particles. A second-order transition was observed with high values of noise ($\eta > 0.2$) and small attraction radius, and a transition compatible with a first-order phase transition was obtained for $\eta < 0.05$ and at longer attraction radius.

Here, we first explore the behavior of the system order-disorder phase transition as a function of v_0 and ρ, for different combinations of η for the original Vicsek model. We fixed the group size to $N = 10,000$ particles, and we plotted the value of $\langle\phi\rangle$ as a function of particles' density on the interval $[0.1, 2]$, and of particles' speed on the interval $[0, 5]$. We wanted to compare the transitions for different noise amplitudes, so we analyzed the changes in $\langle\phi\rangle$ for $\eta = 0.3, 0.4, 0.5, 0.6$, as shown in Figure 4.5.

We found that when both v_0 and ρ are increased, the system undergoes a disorder-to-order phase transition. A dense system and faster moving particles promote more interactions between particles, which in turn endow each particle with a higher number of neighbors to average their direction of motion, and counteract the randomness introduced by the noise term. For this reason, when the noise amplitude is high the system has more difficulties counteracting its effect on particles' direction, and an ordered state is more difficult to achieve.

For low values of η, the disorder-to-order phase transition occurs at very low values of v_0 and ρ. For higher noise amplitudes, the intervals of particle's speed and density where the system is in a disordered state widens. The transition to order occurs at higher values of both v_0 and ρ as η increases. For example, while for $\eta = 0.3$, $\langle\phi\rangle \sim 0.8$ at $\rho = 1.0$ for all v_0, for the case of $\eta = 0.6$, for the same density $\langle\phi\rangle$ would be close to 0 for $v_0 < 1$ and would slowly increase with speed up to only $\langle\phi\rangle \sim 0.3$ for $v_0 = 5$.

From Figure 4.5 we can also observe that the disorder-to-order phase transitions for different critical values have distinct slopes. Focusing on the effect of the particles speed in the system ordering, we explore with more depth the phase transitions as a function of speed. We fixed the particles density to $\rho = 0.2$ where a transition is

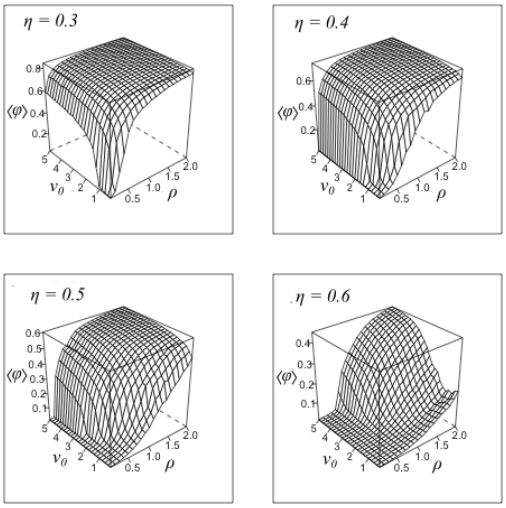


Figure 4.5: $\langle\phi\rangle$ as a function of ρ and v_0 for a fixed value of η. $N = 10,000$. $\langle\phi\rangle$ was averaged over $100,000$ time steps (5,000 time steps of thermalization) in three simulations repetitions.

observed by increasing speed in all $\eta < 0.6$ and we plotted in Figure 4.6A the values of $\langle\phi\rangle$ as a function of v_0 for $\eta = 0.1, 0.2, 0, 3, 0.4, 0.5$ for $N = 10,000$. When the noise amplitude is very low, as is the case of $\eta = 0.1$, the system is ordered even for very slow moving particles. As η increases, the critical v_0 value at which the phase transition is observed becomes higher until noise is too strong ($\eta = 0.5$) for an ordered state to emerge at any v_0 value. Before this occurs, there are η values where the transition from disorder to order seems compatible with a first order transition, as is the case of $\eta = 0.4$. A similar behavior is observed for $\langle\phi\rangle$ as a function of v_0 for different particles' density in Figure 4.6B, where for high density values ($\rho = 1$) the system is fairly ordered for all speed values and completely disordered for low densities values, with medium densities showing a phase transition compatible with a first order transition.

To investigate this further, we chose a value of ρ where a transition was observed at each η fixed value, and examined $\langle\phi\rangle$ as a function of v_0. Figure 4.7 shows the transitions for four (η, ρ) combinations as v_0 changes. For the smaller (η, ρ) values, for range of small values of v_0 , there is little to no increase in the system's $\langle\phi\rangle$, but after a v_0 threshold value, it can be observed that the phase transition seems compatible with a first-order transition.

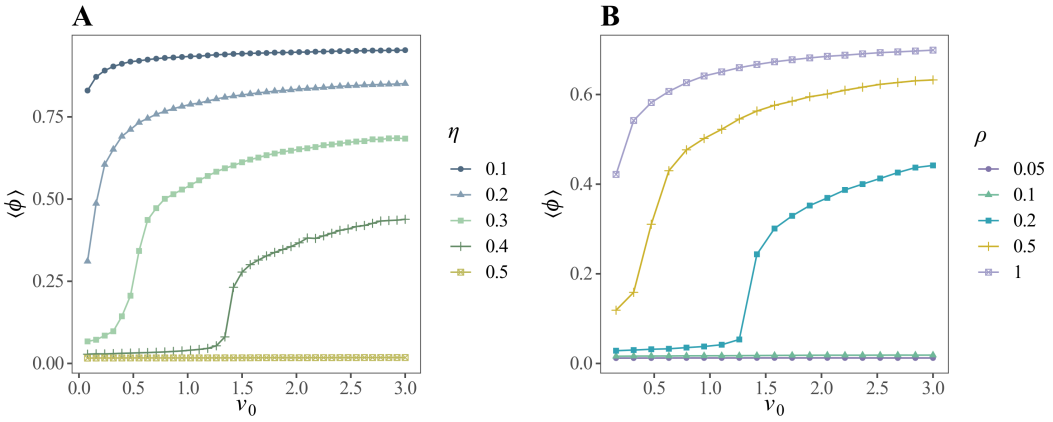


Figure 4.6: **A**: $\langle\Phi\rangle$ per v_0 value, for multiple η values, with fixed $\rho = 0.2$. **B**: $\langle\Phi\rangle$ per v_0 value, for multiple ρ values with fixed $\eta = 0.4$. N = 10,000. $\langle\phi\rangle$ was averaged over 100,000 time steps (5,000 time steps of thermalization) in three simulations repetitions.

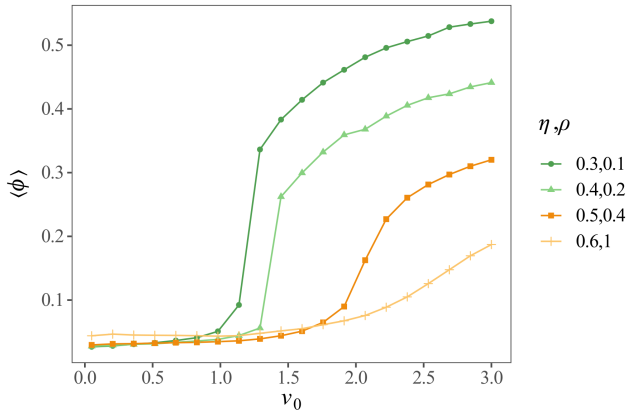


Figure 4.7: $\langle\phi\rangle$ per v_0 value, for combinations of η and ρ where a transition is observed. $N = 10,000$, time = 100,000 (therm = 5,000). Four replicates per η value.

Of course, this assertion needs to be corroborated with a more thorough exploration of the parameter space and the nature of the phase transition, which is outside of the scope of this section. Our objective is to qualitatively show how the Vicsek model's control parameters affect ordering of the system.

Finally, we included in Figure 4.8 the behavior of $\langle\phi\rangle$ as a function of v_0 for parameter combinations that are closer to our experimental observations in fish, i.e. a smaller group size ($N = 100$) and a lower particles' density, $\rho = 0.05$. This ρ value aims to represent a rough approximation of fish density inside the experimental tank, calculated as,

$$\rho = \frac{Nl^2}{L^2} \tag{4.4}$$

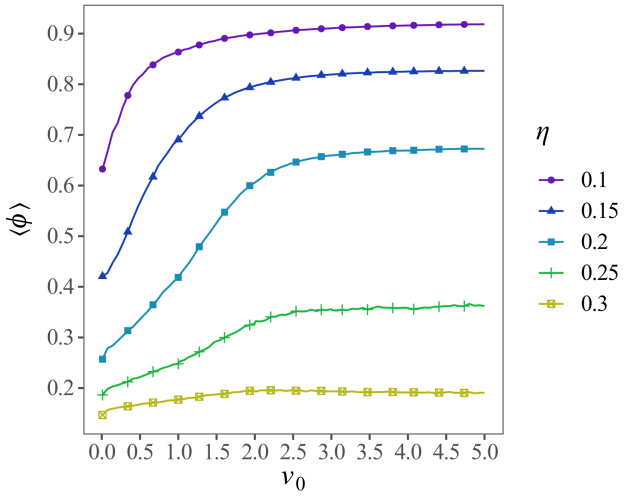


Figure 4.8: $\langle\Phi\rangle$ per v_0 value, for multiple η values. $N = 100$ $\rho = 0.2$. $\langle\phi\rangle$ was averaged over 100,000 time steps (5,000 time steps of thermalization) in four simulations repetitions.

assuming $l = 2$ cm and $L = 100$ cm. In Chapter 5 we compare the order dependence of speed in fish with a modified version of the Vicsek model. The disorder-to-order transition was also observed for this parameter combination under a noise amplitude threshold ($\eta < 0.25$). Yet, the transition is smoothed out by the low density of particles. We can also observe $\langle\phi\rangle$'s saturation values depending on the η value for threshold values of v_0 , where $\langle\phi\rangle$ no longer increases.

The disorder-to-order phase transition that we observed as particles' speed is varied differs from the conclusions in Cambui and Rosas, 2018. Our results point to a disorder-order phase transition as speed increases, while in their study the transition they found was from order-to-disorder. This disparity can stem from the effects of the attraction radius on the collective ordering of the system that Cambui and Rosas, 2018 included in the model. We worked with the original Vicsek model, which only considers an interaction radius where particles align with their neighbors. As it will be mentioned in Chapter 5 the implementation of an attraction zone can provoke the alignment forces to lose correlation with speed.

The simplicity of the Vicsek model suggests that phase transitions are a possible universal feature of moving groups. From a biological perspective, this observation is of great relevance, as it suggests that phase transitions could be an intrinsic consequence of grouping, and not a natural selection fine tuning (Sumpter et al., 2008).

5 Collective ordering dependence on swimming speed

5.1 Introduction

Fish ensembles will often align their headings through collective movement as a means to travel together. Multiple experimental studies have observed a correlation between group directional synchronization and the swimming speed of individuals. In various species, there is a demonstrable increase in collective order as individuals increase their swimming speed (Viscido et al., 2004; Gautrais et al., 2012; Tunstrøm et al., 2013; Rieucau et al., 2015).

From an adaptive perspective, it is understandable that speed acts as a modulator of collective ordering. Responding to speed changes requires minimum perception and cognitive abilities, and functions as a mechanism to transfer information quickly and efficiently (Lemasson et al., 2013), allowing groups to react rapidly to changing environmental conditions. In the pursuit of linking local behavioral rules with emergent collective states, it is imperative to understand the dependence of the collective order of a flock on the speed of the individuals.

Inspired by experimental observations of fish, various models have been proposed to infer the local interaction rules that govern group cohesion and directional synchronization. These data-driven models include either active heading synchronization dependency to a particle's speed, or this correlation emerges as a by-product of the model's behavioral rules. However, it is still not clear if this dependence can emerge in a simpler model of collective motion such as the Vicsek model, described in Chapter 4. Modified versions of the Vicsek model that have studied the effects of speed in particles' collective ordering either include other behavioral rules

that change the collective dynamics, or actively force the polarization-speed relationship. These model modifications might not be necessary to reproduce experimental observations of collective ordering dependence on swimming speed.

Next, we review data-driven and Vicsek based models that address this behavior, highlighting their findings in speed-polarization correlation.

5.1.1 Collective ordering and speed in models of collective motion

Data-driven models

Viscido et al., 2004 first quantified the relationship between polarization and speed by analyzing schools of four and eight giant danios (*Danio aequipinnatus*. By characterizing the individuals movements, they constructed an agent-based model centered on attraction/repulsion and alignment forces, and concluded that speed was mediated mainly by the alignment force, and was positively correlated with polarization. Couzin et al., 2002 also proposed a collective motion model based on repulsion, attraction and alignment of orientations, and showed that transitions from disorganized to organized collective states primarily involve mainly the adjustment of individual speeds. Gautrais et al., 2012 built a model based on the quantification of individual motion and local interactions in groups of barrel flagtails (*Kuhlia mugil*). They found that alignment and attraction interactions were able to replicate the motion behavior, which in turn were regulated by speed. This feature was further investigated by Calovi et al., 2014, which found the existence of two dynamically stable collective states: a swarming state in which individuals aggregate without cohesion, with a low level of polarization, and a schooling state in which individuals are aligned with each other and have a high level of polarization. They concluded that mean speed acts as a modulator on the strength of each interaction. High speeds mean the predominance of alignment interactions over attraction.

Herbert-Read et al., 2011 analyzed groups of two to eight mosquitofish (*Gambusia holbrooki*), and created a model through machine learning techinques to infer the interaction rules individual fish follow to maintain group structure. They found that attraction forces are responsable for maintaining group cohesion, but contrary to previous studies, they did not find evidence that fish actively match orientation

with their neighbors. Instead, fish actively changed their speed in order to avoid or move toward neighbors, i.e. attraction/repulsion forces were regulated by speed, which in turn could modulate group alignment. However, they did not discard the possibility of species changing their adopted rules depending on context.

Tunstrøm et al., 2013 studied the emergence of collective states in groups of up to 300 individuals of golden shiners (*Notemigonus crysoleucas*), as well as their dependence on individual responses and on external perturbations. They found three dynamically-stable collective behaviors: swarm (disordered state), polarized (aligned state) and milling (fish moving around the center of the group; examples can be seen in Fig.2.1 in Chapter 2), which correlated to group size and individual speed. Low speed was associated with the group being dense and disordered, i.e. swarming, and high speed was associated with polarized or milling groups.

From a kinematic perspective, Ashraf et al., 2016 studied tail beatings synchronization of pairs and triads of red nose tetra fish (*Hemmigrammus bleheri*), a close relative to the fish we use in our study (black neon tetra described in Chapter 2). Ashraf et al., 2016 characterized the time fish spent synchronized, either out of phase (tails to opposite side) or in phase (tails to same side), and non-synchronized, for all swimming velocities. They found that for relatively small swimming speeds, fish spend a majority of their time swimming independently in a non-synchronized state. Increasing swimming speed changes this tendency to synchronized states until fish are fully synchronized at higher velocities. Regardless of the speed, distance between neighbors remains constant. This is an indication that speed can be viewed as a modulator of the strength og the interaction between individuals, as hydrodynamic pressure between neighbors increases with speed.

The Vicsek model (Vicsek et al., 1995; see Chapter 4) has been widely used as an archetypical model to explore possible universalities within the study of collective motion. The model predicts the presence of a phase transition separating a disordered, swarming phase, from an ordered, flocking one by modulating noise strength (Vicsek et al., 1995; Ginelli, 2016). The relation between a variable speed and the order-disorder phase transition in the Vicsek model and in variations of it has been considered in several works. For example, Cambui and Rosas, 2018 considered a variable speed in a modified Vicsek model with an attraction rule to mantain

the cohesion of particles. Contrary to what has been documented in experimental observations and in data-driven models, they found that increasing average speed decreased the order of the system, and that the nature of the order-disorder phase transition was dependent on noise intensity and interaction radius between particles. Mishra et al., 2012, on the other hand, coupled the local order with the local speed to force the system to improve local polarization at local high speed, following empirical observations in real fish. However, this model does not explain the intrinsic dependence of the two parameters. Contrary to the observed behavior of Cambui and Rosas, 2018 in the Vicsek model, our exploration of the original model in Chapter 4 shows that the collective order undergoes a phase transition from order to disorder depending on the speed of the particles. This straightforward model appears to reproduce the findings of studies on actual schools of fish.

Finally, an example of how swimming speed impacts the ordering of fish collectives is described in Herbert-Read et al., 2015. This study investigated pacific blue-eyes (*P. signifer*) and found that changes in speed by a few informed individuals can initiate escape waves. They modeled the dynamics observed by modifying the Vicsek model to include particles' acceleration and deceleration, and confirmed that the group probabilities of changing direction increase when informed individuals increase their speed. They concluded that escape waves occur when there is an initial increase in an individual's speed that leads to a local density increase that propagates through the group.

5.1.2 Objectives

In this chapter we aim at contributing to the understanding of collective order dependence upon speed by studying within an experimental setting and utilizing a variation of the Vicsek model, in order to see whether such a simple model is able to account for this dependence.

From the experimental point of view, we study the relationship between swimming speed and polarization in groups of black neon tetra (*Hyphessobrycon herbertaxelrodi*), a small freshwater fish that has a strong tendency to aggregate in polarized schools, even when habituated to a novel environment (Gimeno et al., 2016, See Chapter 2).

We observed locomotion in our experimental subjects that consisted in alternating bouts of active swimming, with passive coast periods, a strategy called burst-and-coast.To our knowledge, this behavior is not formally characterized specifically in black neon tetra, but is a motion behavior characteristic of body-caudal fin swimming modes (BCF), which represents the fish used in our study. This movement includesthat include body and/or caudal fin oscillations as a way of propelling thrust to increase swimming speed (Webb, 1984). This swimming strategy has been documented in multiple fish species (Korsmeyer et al., 2002; Fish et al., 1991; Wu et al., 2007; Dutil et al., 2007), including other tetra fish pertaining to the same family as our study subjects (Li et al., 2021; Calovi et al., 2018).

For the theoretical analysis, to imitate the speed behavior observed in black neon tetra, we modified the Vicsek model so that particles' speed oscillates as a triangular wave in time with a characteristic period, and analyzed changes in polarization as a consequence.

5.2 Results

5.2.1 Experimental data

To quantify fish speed and collective ordering we categorized the motion of groups of $N = 20$ and $N = 40$ individuals of black neon tetra (*Hyphessobrycon herbertaxelrodi*) swimming freely in an experimental tank. Fish (x,y) coordinates in time were obtained for three replicates of 12,000 frames each (ABC) for each group size, adding up to a total of 72,000 frames analyzed ($\approx$ 1 hour). Group size conditions, tagged as Q5N20 for $N = 20$ and Q5N40 for $N = 40$, were analyzed separately to avoid mixing potential effects of differing group size. Physical units were converted from pixels/frames to centimeters/seconds using the camera frame rate and resolution, and the tank dimensions (see Chapter 2).

From the sequence of positions of fish in time, the discretized instantaneous velocity was computed using a Richardson extrapolation scheme of order 4 (Fornberg, 1988), in terms of the expression

$$\vec{v}_i(t) = \frac{1}{12\Delta t}\left[\vec{r}_i(t-2) - 8\vec{r}_i(t-1) + 8\vec{r}_i(t+1) - \vec{r}_i(t+2)\right], \tag{5.1}$$

where $\Delta t = 1/20$s the time discretization step.

From $\vec{v}_i(t)$, we obtained the instantaneous speed of each fish,

$$v_i(t) = |\vec{v}_i(t)| \equiv [\vec{v}_i(t) \cdot \vec{v}_i(t)]^{1/2}, \tag{5.2}$$

the average speed of the flock,

$$\bar{v}(t) = \frac{1}{N} \sum_i v_i(t), \tag{5.3}$$

and the instantaneous polarization,

$$\phi(t) = \frac{1}{N} \left| \sum_{i=1}^{N} \frac{\vec{v}_i(t)}{v_i(t)} \right|, \tag{5.4}$$

measuring the global degree of order in the flock at time t (Vicsek et al., 1995; Ginelli, 2016). In a perfectly ordered flock, in which all fish travel in the same direction, albeit with possibly different speeds, ϕ takes its maximum value 1. In a disordered swarm, with all fish traveling in randomly chosen directions, one expects to observe a very small average value, scaling as $\langle \phi \rangle \simeq N^{-1/2}$ (Ginelli, 2016).

Speed statistics

The probability density functions (PDF) for both instantaneous $P(v_i)$ and average speed $P(\bar{v})$ for both conditions (each consisting of replicates A,B and C) are shown in Figure 5.1. Both distributions are bell shaped, lacking long tails, with a mean value 18.4 ± 7.3cm/s. The distribution is additionally stationary in time, with non overlapping slices of the time sequence giving very similar distributions. For example, the middle half of the sequence of individual velocities has a mean value 18.3 ± 6.8cm for condition Q5N40, fully compatible with the whole sequence. Mean values and error bars for both analyzed conditions are summarized in Table 5.1. Both $\langle v_i \rangle$, $\langle \bar{v} \rangle$ and r_{v_i,v_j} are slightly lower for groups of 20 individuals.

In Figure 5.2(A) we plot the individual speed as a function of time for a sample of 3 randomly chosen fish from condition Q5N40. As we can see from this figure, black

Table 5.1: Mean $\pm S.D.$ individual and average swimming speeds in time, and mean Pearson correlation coefficient for individual speeds, original and randomized.

Condition	$\langle v_i \rangle$ **(cm/s)**	$\langle \bar{v} \rangle$ **(cm/s)**	r_{v_i,v_j}	$r_{v_{irand},v_{jrand}}$
Q5N40	18.4 ± 7.3	18.4 ± 6.8	0.51 ± 0.11	0.0002 ± 0.03
Q5N20	17.2 ± 8.5	17.2 ± 5.96	0.48 ± 0.09	0.0003 ± 0.03

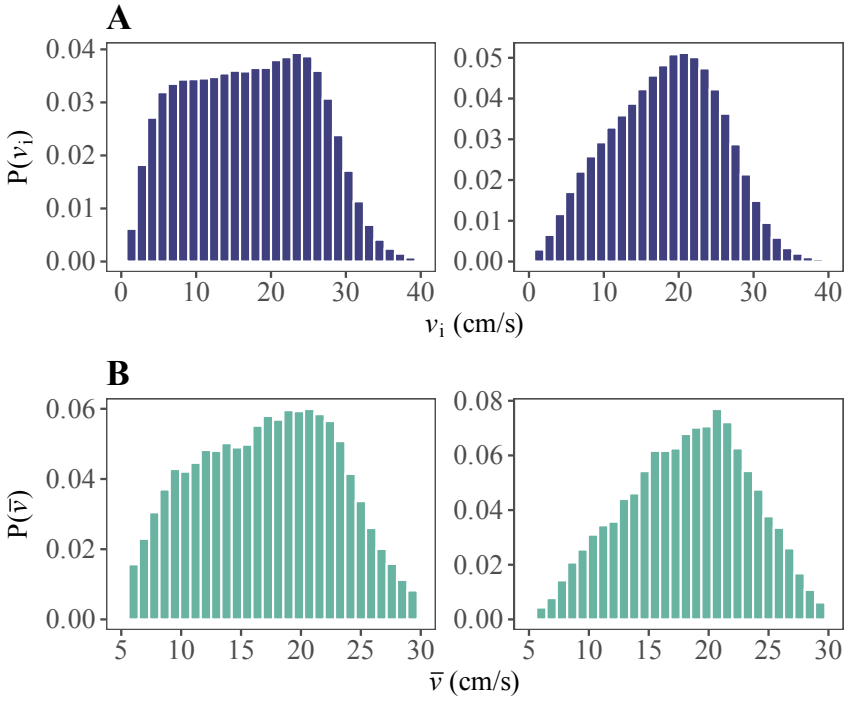


Figure 5.1: **A**: Probability density function (PDF) $P(v_i)$ of the individual instantaneous speed v_i. **B**: PDF $P(\bar{v})$ of the average instantaneous speed $\bar{v}$.Left, condition Q5N20. Right, condition Q5N40.

neon tetra's individual speeds show a clear quasi-periodic oscillation pattern that is consistent with a burst-and-coast behavior (Fish, 2010a), where we can interpret the regions of increasing speed as corresponding to the burst phase, while the periods where the speed decreases represent the result of the coast phase.

Moreover, the oscillation of speed over time seems to be synchronized between the different individuals, indicating that the beginning of the acceleration (dips) and gliding (peaks) phases in the different fish take place approximately at the same instant of time, with a potential small phase shift. This observation suggests a synchronized individual speed behavior reflected in collective swimming, which has been already observed in red nose tetra fish (*Hemigrammus bleheri*), that synchronize

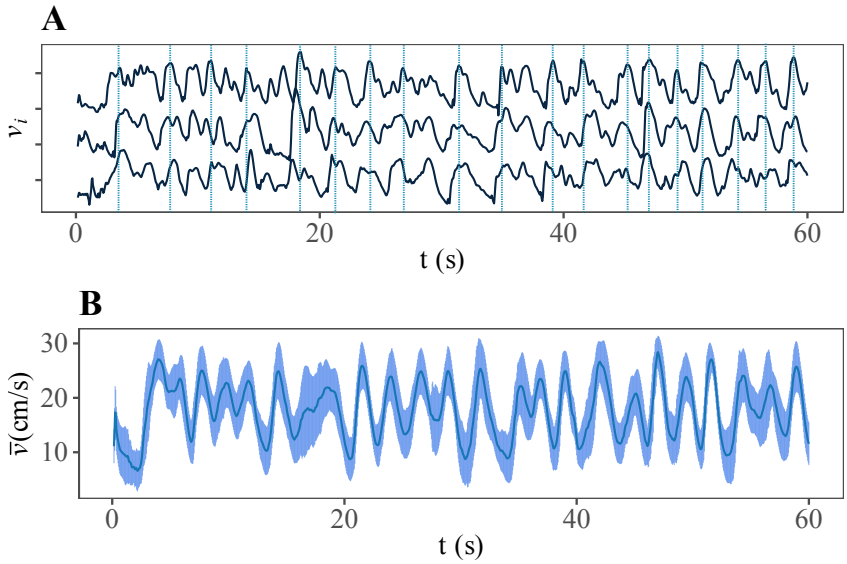


Figure 5.2: A: Individual speed $v_i(t)$ as a function of time for three randomly selected fish in a 60 seconds time lapse in replicate B in Q5N40. Dotted vertical lines mark selected local maxima of the upper curve. The plots have been shifted vertically for clarity. **B**: Fish average speed $\bar{v}$ as a function of time for same time laps in replicate B in Q5N40. The shaded region correspond to one standard deviation.

their burst-and-coast events and save energy by limiting water flow speed fluctuations (Calovi et al., 2018). In Figure 5.2(B), we show the evolution in time of the average speed of the flock $\bar{v}(t)$. Here we can observe that the average speed also experiences an oscillatory evolution, as expected from the synchronization of individual speeds, and reflected in the small dispersion measured by the average speed standard deviation, shown in the shaded region in Figure 5.2(B).

A more quantitative characterization of individual speed synchronization can be obtained by evaluating the Pearson correlation coefficient r_{v_i,v_j} of the sequence of instantaneous velocities of pairs of fish. In Figure 5.3(A) we show the PDF $P(r_{v_i,v_j})$ corresponding to the Pearson correlation of pairs of actual velocity sequences for both experimental conditions (mean values are reported in Table 5.1). In order to increase the statistics, we computed the Pearson coefficient for each replicate over 10 non-overlapping windows of length 1200 frames (60s). Moreover, to confirm that fish (r_{v_i,v_j}) values were not merely a product of randomness, we randomized the individual speed sequences by sampling values from v_i distribution, and calculated (r_{v_i,v_j}).

In the randomized case, correlations are essentially nonexistent, with an almost zero mean Pearson coefficient $r_{v_i,v_j\,\mathrm{rand}} = 0.0002 \pm 0.03$ and a maximum value 0.107;

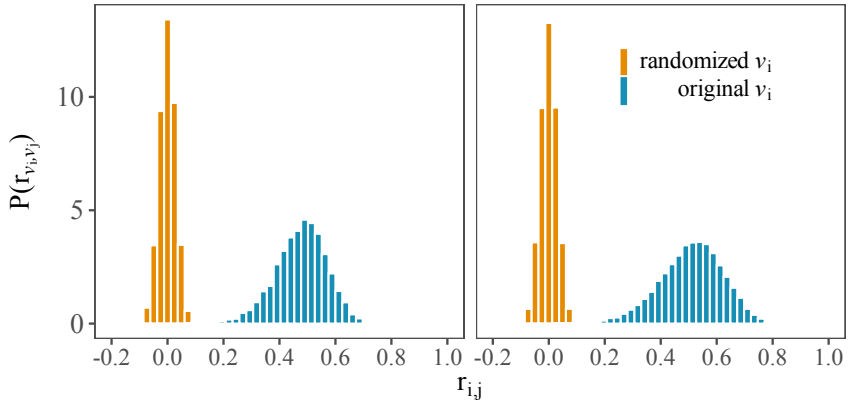


Figure 5.3: PDF $P(r_{ij})$ of the Pearson correlation coefficients $r_{i,v}$ of the instantaneous speed of pairs of fish (blue). The orange curve represents the PDF of Pearson coefficients obtained for randomized sequences of individual speeds. Left, condition Q5N20. Right, condition Q5N40.

in our experimental fish, however, large Pearson coefficients are observed, with an mean value $r_{v_i,v_j} = 0.51 \pm 0.11$, and a maximum value up to 0.84 for condition Q5N40, indicating that the behavior in time of the speed of pairs of fish is strongly correlated.

Correlation between speed and collective ordering

In order to investigate the effects of a quasi-periodic oscillation speed on the collective order of black neon tetra, we have computed the instantaneous polarization of the flock $\phi(t)$, defined in Eq.(5.4). In Fig.5.4(A) we plot a sample of $\phi(t)$ as a function of time for a segment of 60s of replica B of condition Q5N40. As we can see, flocks show a high level of polar order (see Table 5.2), with a mean value higher than 0.92 for both group sizes, remaining highly polarized most of the time, as reflected by the small standard deviation. These values are in agreement with the results reported in Gimeno et al., 2016 for black neon tetras. The evolution of the polarization in time shows also an apparent oscillatory behavior, similar to the average speed. Indeed, in Figure 5.4(B), where we present a simultaneous standardized plot of average velocity and polarization, we can see that polarization tends to follow the behavior of the mean speed, with almost coincident relative maxima and minima.

These correlations are also evident in a color density plot of the polarization $\phi(t)$ as a function of the average speed $\bar{v}(t)$ in Fig.5.5. Again, this plot shows the strong

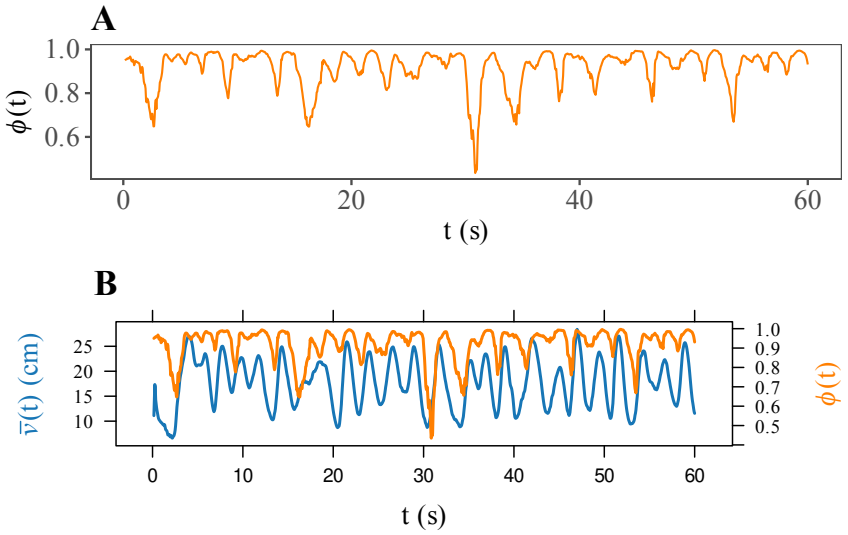


Figure 5.4: **A**: Instantaneous polarization $\phi(t)$ as a function of time. **B**: Standardized plot of the average speed $\bar{v}(t)$(blue) and polarization $\phi(t)$ (orange) as a function of time. Condition Q5N40, replicate B.

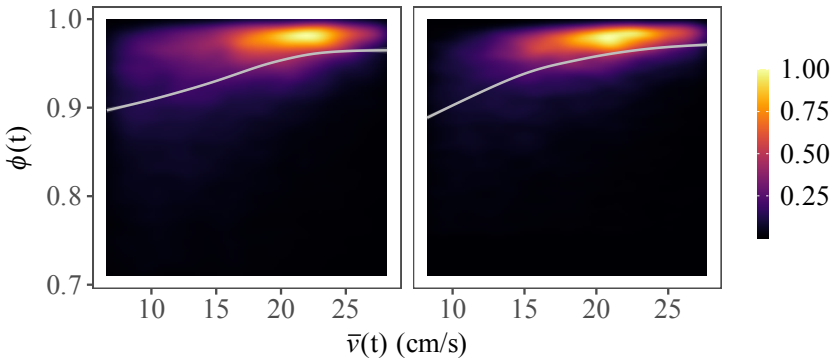


Figure 5.5: Normalized color density plot of the polarization $\phi(t)$ as a function of the average velocity $\bar{v}(t)$ for all replicates of each condition.Left, condition Q5N20. Right, condition Q5N40. White lines represent the mean polarization as a function of average speed.

correlations between the evolution of the speed and the polarization, with polarization taking larger values when the speed increases, and smaller values when the speed decreases. This is particularly evident in the plot of the mean temporal polarization as a function of the average speed, depicted as a white line in Figure 5.5. Our observations agree with independent experiments of collective motion in fish reported in the literature (Gautrais et al., 2012; Tunstrøm et al., 2013; Rieucau et al., 2015; Viscido et al., 2004).

To measure the correlation between polarization and speed, and to account for

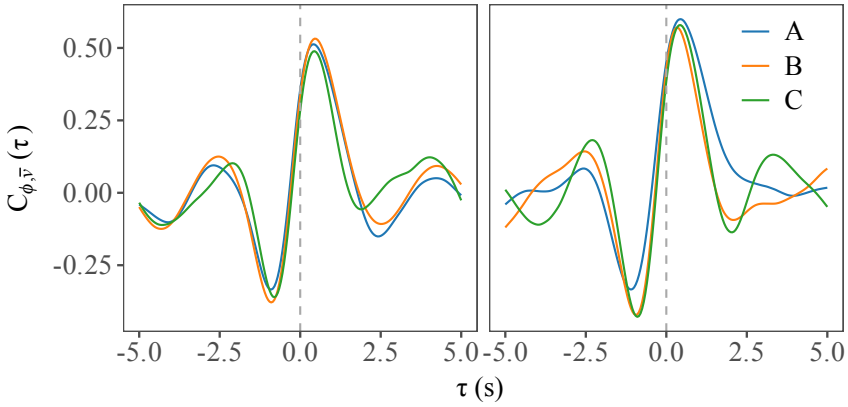


Figure 5.6: Normalized cross-correlation $C_{\phi,\bar{v}}(\tau)$ of polarization and average velocity as a function of the time lag τ for the three replicates of each experiment. Left, condition Q5N20. Right, condition Q5N40. Cross-correlations are computed over a window of 60s

.

the small shift we observe in these parameters time series maxima and minima, we computed the normalized cross-correlation (CCV),

$$_{\phi,\bar{v}}(\tau) = \frac{1}{t_{max}} \sum_{t=1}^{t_{max}} \frac{(\phi(t) - \phi)(\bar{v}(t+\tau) - \bar{v})}{\sqrt{\sigma_\phi \sigma_{\bar{v}}}}, \tag{5.5}$$

where ϕ ($\bar{v}$) and σ_ϕ ($\sigma_{\bar{v}}$) denote mean and standard deviation of the polarization (resp. average speed). Same as in individual speeds Pearson correlations, to increase statistics we computed correlations over 10 non-overlapping time windows of size 1200 frames.

From this analysis, see Figure 5.6 and Table 5.2, we observe that the conditions' instantaneous correlation at time lag $\tau = 0$ is higher than 0.4 for both conditions. However, the maximum correlation between average speed and polarization takes place for a lag $\tau_{\max}$ larger than zero, that we estimate as approximately eight frames for both conditions analyzed, corresponding to a mean maximum correlation $C_{\phi,\bar{v}}(\tau_{\max}) > 0.5 C_{\phi,\bar{v}}$ for both τ_0 and τ_{max} were lower for Q5N20. When this lag is taken into account, we observe that the correlations between $\bar{v}(t)$ and $\phi(t + \tau_{\max})$ become better for both conditions.

As we can see in the standardized plot in Fig.5.7(A), the maxima and minima of $\bar{v}$ in time correspond more accurately with the maxima and minima of $\phi(t + \tau_{\max})$. On the other hand, in the color density plot of $\bar{v}$ versus $\phi(t + \tau_{\max})$, Figure 5.7(B),

medium and low values of ϕ are no longer observed at higher $\bar{v}$: the system's order is low only when average fish speed is sufficiently low.

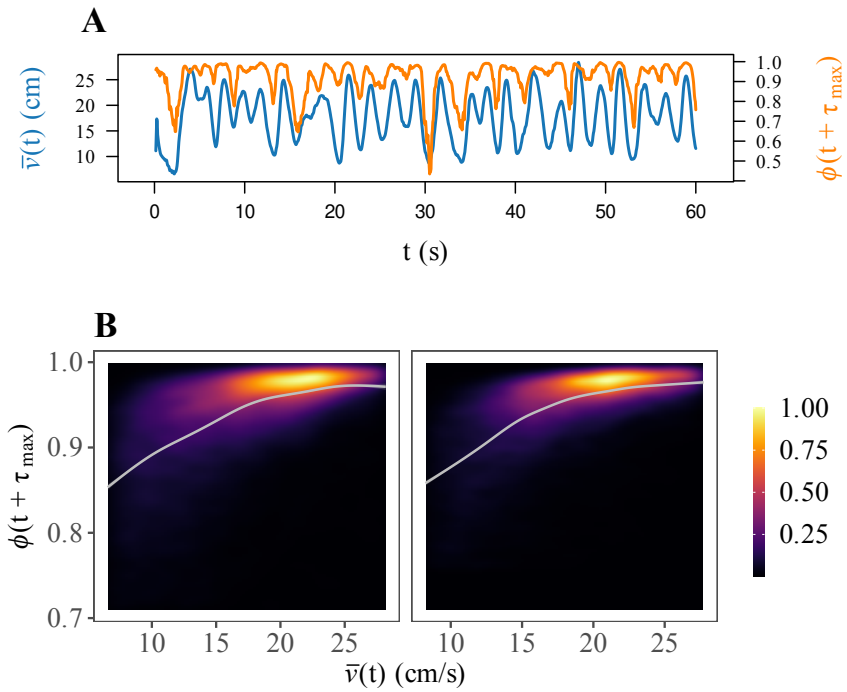


Figure 5.7: **A**: Standardized plot of the average speed $\bar{v}(t)$(blue) and shifted polarization $\phi(t + \tau_{\max})$ (orange) as a function of time for condition Q5N40, where $\tau_{\max}$ is the lag that maximizes the normalized cross-correlation. **B**: Normalized color density plot of the shifted polarization $\phi(t + \tau_{\max})$ as a function of the average velocity $\bar{v}(t)$ for all replicates of each experiment. Left, condition Q5N20. Right, condition Q5N40. White lines represent the mean polarization as a function of average speed.

Such non-zero lag might imply the presence of a characteristic response time between velocity and polarization fluctuations. Indeed, it is plausible that, after a sudden reorientation of the flock, and therefore inducing a change of velocity, it might take some time for the fish in the flock to adjust their trajectories to return again to their level of optimal polarization.

Table 5.2: Mean polarization in time and mean normalized cross-correlation values at τ_0 and τ_{max} for swimming speed and polarization.

Condition	$\langle\phi\rangle$	$C_{\phi,\bar{v}}(\tau_0)$	$C_{\phi,\bar{v}}(\tau_{\max})$	τ_{max} **(s)**
Q5N40	0.93 ± 0.01	0.42 ± 0.04	0.58 ± 0.01	0.42 ± 0.06
Q5N20	0.92 ± 0.11	0.33 ± 0.04	0.51 ± 0.02	0.43 ± 0.03

Table 5.3: Mean oscillation periods T for individual speed (s), average speed (s), polarization, and individual and centroid distance to border (s).

Condition	T_{v_i}	$T_{\bar{v}}$	T_{ϕ}	T_{ID}	T_{CD}
Q5N40	1.90 ± 0.87	2.68 ± 1.05	2.40 ± 0.96	6.00 ± 1.96	6.24 ± 1.71
Q5N20	1.66 ± 0.75	2.45 ± 0.97	2.43 ± 0.97	5.50 ± 1.55	5.65 ± 1.42

Quasi-periodic nature of velocity and polarization

As we have noted above, both velocity and polarization over time seem to show a quasi-periodic behavior, which in the case of the speed we can attribute to the burst-and-coast swimming mechanism of the fish. In order to explore in more detail this periodicity, we attempted to determine the value of the average period on both speed and polarization. The short length of our experimental time series precludes the application of a standard Fourier analysis. Therefore, we opt for a more empirical approach, measuring the distance between consecutive maxima in the time series $\phi(t)$, $v_i(t)$, and $\bar{v}(t)$, after applying a smoothing one-dimensional Gaussian filter of window length 1s for both speeds and 3s for polarization.

In Figure 5.8(A) we plot the PDF $P(T)$ of periods thus computed for the individual velocity, the average flock velocity and the instantaneous polarization. These observed distributions exhibit a strong overlap, specially for Q5N40, and are characterized by similar mean periods.A summary of the statistics of the empirical periods of speed and polarization is presented in Figure 5.8(B) in the form of violin plots, and mean values documented in Table 5.3. From here we can conclude that the empirical periods of the average speed and polarization are very strongly correlated, being equal, within error bars, at the level of the mean values.

Effects of the walls

It could be argued that the oscillations in the swimming speed of the fish could have their origin in the boundary effects of the walls of the experimental tank, which force the fish to slow down and change their trajectories when they approach them. These boundary effects would impose a quasi-periodic variation of the speed, with a period related to the average time to cross the tank between opposite walls, and would therefore not be related to any self-organized synchronization process or to

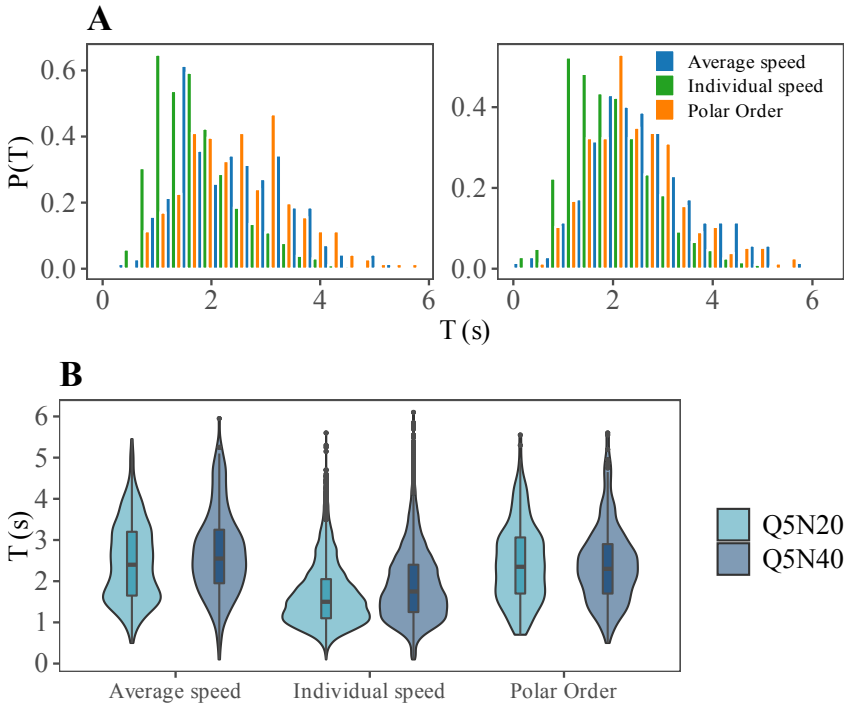


Figure 5.8: **A**: PDF $P(T)$ of empirical periods, defined as thetemporal distance between consecutive maxima, for the average speed, individual speed and instantaneous polarization. Left, condition Q5N20. Right, condition Q5N40. **B**: Violin plot summarizing the statistics of the distributions of speed and polarization empirical periods. Data refers to the aggregation of all replicates for both conditions.

the burst-and-coast swimming mechanism. In consideration of this possibility, we have studied the correlations between the speed and the distance to the tank walls.

In particular, we have determined two different correlations: The correlation between the average speed $\bar{v}(t)$ with the distance of the centroid of the flock to the closest wall of the tank, and correlation between the instantaneous speed $v_i(t)$ of each individual fish with the distance of the fish to the closest wall. These correlations are analyzed in the form of normalized color density plots in Figures 5.9(A) and (B). The results demonstrate a weak correlation between velocity and wall distance, for either the flock as a whole or at the individual fish level. This observation suggests that although the tank walls may have an effect on the swimming dynamics of the fish, they are not the primary force ruling the quasi-periodic variation of the individual and average speeds. Therefore this behavior can be more appropriately attributed to the burst-and-coast strategy.

Interestingly, both the individual and centroid minimum distances to the walls also show a quasi-periodic dynamics, although this is not as well indicated as in the

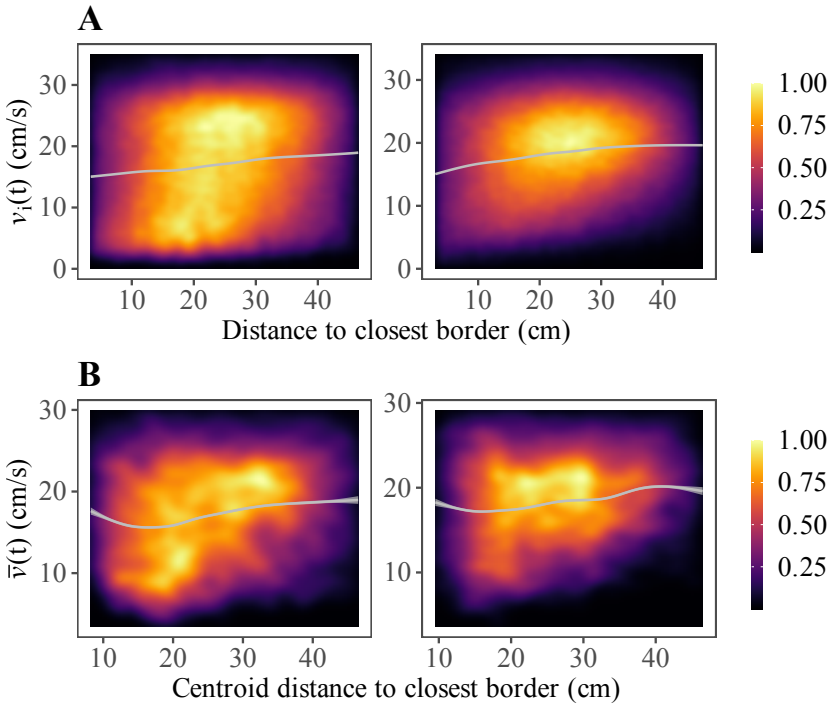


Figure 5.9: **A**: Normalized color density plot of the individual speed $v_i(t)$ as a function to the distance of the fish to the closest tank wall. **B**: Normalized color density plot of the average school speed $\bar{v}(t)$ as a function to the distance of the school centroid to the closest tank wall. Left, condition Q5N20. Right, condition Q5N40.

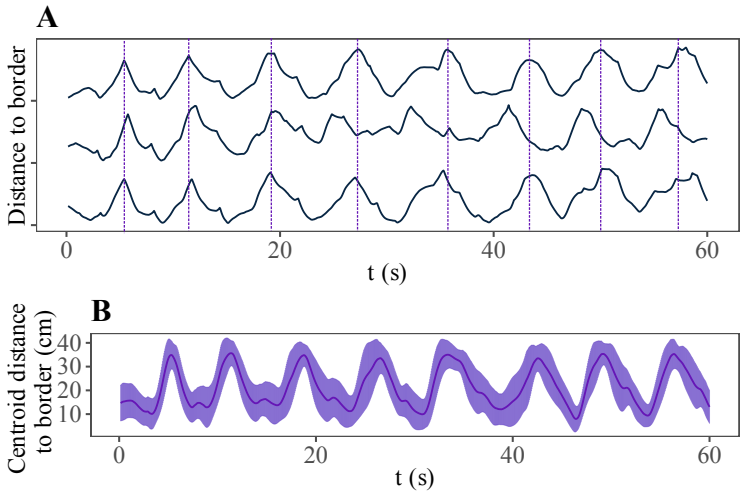


Figure 5.10: **A**: Individual distance to nearest tank border as a function of time for three randomly selected fish in a 60 seconds time lapse in replicate B of condition Q5N40. Dotted vertical lines mark selected local maxima of the individual in the upper curve. The plots have been shifted vertically for clarity. **B**: Fish centroid distance to nearest tank wall as a function of time for replicate B of Q5N40. The shaded region corresponds to one standard deviation.

case of speed and polarization, see Figures 5.10(A) and (B), specially for the case of individual distances.

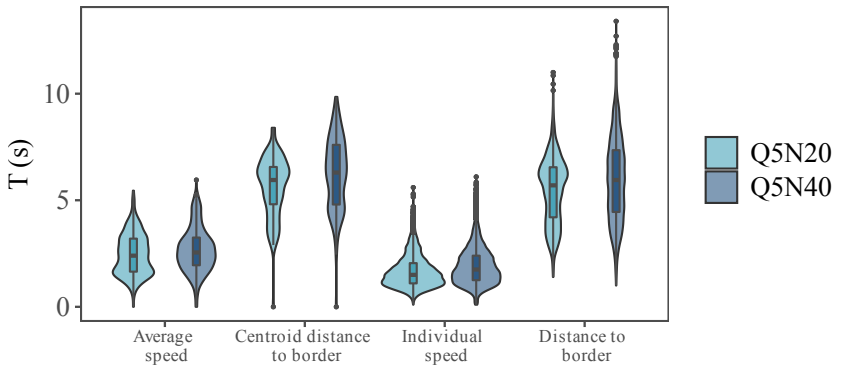


Figure 5.11: Violin plots summarizing the statistics of the distributions of empirical periods of the distance to the closest wall of individual fish and the school centroid, compared with the distributions of individual and average speeds. Data refers to the aggregation of all replicates for both experimental conditions

We have estimated the empirical periods of these quantities, following the approach used for the case of speed and polarization. In this case, we used a smoothing Gaussian filter of window length 4s. In Figure 5.11(A) we summarize the statistics of the distributions of empirical periods measured, in terms of box plots, compared with the distributions of average and individual speed. As we can observe, the distributions of periods to the distances to the walls are broader than the ones corresponding to speed, taking mean values $T_{\mathrm{CD}} = 6.3 \pm 2.6$ and $T_{\mathrm{ID}} = 5.9 \pm 1.9$ for the wall distance of the centroid and individuals, respectively, and practically equal within error bars. However, these periods are more than double the periods observed for speed and polarization in both conditions, which further strengthens the conclusion that the quasi-periodic nature of the movement and order of the flock is not induced by interactions with the tank walls.

5.2.2 Modeling speed variation in flocking dynamics

In order to explore in a more controlled way the effects of synchronized individual speed in the ordering dynamics of collective motion, we study a variation of the classical Vicsek model (Vicsek et al., 1995) with varying particle speed. In the original Vicsek model (see Chapter 4), a set of N self-propelled particles travel at constant speed v_0 with an orientational dynamics, thus making their particles align their velocity. THey do this with the mean velocity of a set of neighboring particles inside a circle or radius R_0, and subject to a noise perturbation of amplitude η, accounting

for errors in the determination of the neighbors' state. In our variation of the model, and in order to mimic the behavior observed empirically in black neon tetras, we impose to all particles a periodic time varying speed $v_i(t) = v(t), \forall i$, that, inspired by the burst-and-coast swimming mechanism, and is chosen as a triangular wave of period T, i.e. $v_i(t+T) = v_i(T)$, with a minimum value $v_{\min}$ and a maximum value $v_{\max}$, corresponding to an amplitude $A = (v_{\max} - v_{\min})/2$, given by the expression in the interval $[0, T]$

$$v(t) = v_{\min} + (2A - v_{\min}) \left[1 - \frac{\left| t - \frac{T}{2} \right|}{\frac{T}{2}} \right] \tag{5.6}$$

In order to use a system model compatible with our experimental setting, we consider a number of particles $N = 100$ and a radius of interaction $R_0 = 1$. The density ρ can be estimated from the experimental data assuming that the fish have an excluded surface given by disk of radius ℓ, where ℓ is the fish body length. So we have

$$\rho_{\text{exp}} \simeq \frac{N_{\text{exp}} \ell^2}{L_{2\text{exp}}.} \tag{5.7}$$

With $N_{\text{exp}} = 40$, $L_{\text{exp}} = 100$cm and $\ell = 2.5$cm, we obtain $\rho_{\text{exp}} \simeq 0.07$. In our simulations we thus fix a density $\rho = 0.05$. Statistical properties are computed averaging over 500000 time steps, after discarding an initial transient of 10000 time steps.

When we reviewed the behavior of collective ordering in the Vicsek model in Chapter 4, we saw that the mean polarization ϕ, defined as the temporal average of the instantaneous polarization in Equation (5.4), is an increasing function of v_0 (see Figure 4.8 in Chapter 4). This relationship is in agreement with the average positive correlations between speed and polarization experimentally observed, but at odds with previous numerical simulations (Baglietto and Albano, 2009b; Cambui and Rosas, 2018); we notice, however, that (Baglietto and Albano, 2009b) considers a much smaller range of velocity values and that (Cambui and Rosas, 2018) introduces an additional attraction rule. To fix the parameters of the oscillating speed in our model variant, we analyzed the behavior in the classic Vicsek model of the average ϕ in time, $\langle \phi \rangle$, as a function of v_0 for different noise amplitudes. As we can observe

in this Figure, for $\eta < 0.3$ there is a significant gap in the polarization in the ordered and disordered states.

Fixing a low noise value $\eta = 0.1$, compatible with the highly polarized state shown by black neon tetras, we consider a range of speed values in the interval $[0.01, 2.99]$ covering a region of high and low order, see Figure 5.12(A). We therefore choose as the time varying speed in our model a triangle wave with minimum value $v_{\text{min}} = 0.01$, maximum value $v_{\text{max}} = 2.99$, corresponding to an amplitude $A = 1.49$, and we tested oscillation periods T that covered from $T = 100$ to $T = 3000$.

With this standardized, synchronized periodic speed imposed on all particles in the model, we obtain time series of the instantaneous polarization that also vary in time in an almost periodic form, see Figure 5.12(B). For small period $T = 100$ the polarization oscillates primarly around high values between 0.6 and 0.9. Increasing the period results in the occasional presence of highly disordered states of low polarization close to zero, much like the experimental polarization $\phi(t)$ observed in fish. This is due to the relaxation time needed for the system to accommodate the value of the polarization expected for each temporal value of the time-varying speed. For small periods, velocity changes too quickly, and the model is not able to relax to the disordered state corresponding to small velocity, and thus polarization oscillates primarly among high values. For large periods, instead, the relaxation time is short enough and the system can explore more freely the disordered state.

This relaxation time is comparable to the response time between velocity and polarization variations discussed for real fish, and we can analyze it analogously computing the normalized cross-correlation $C_{\phi,\bar{v}}(\tau)$ between the numerical polarization and the triangular wave speed. For $\tau = 0$, the instantaneous correlation is around 0.4 for short periods, $T \sim 100$, and saturates at a higher value around 0.6 for sufficiently large periods $T > 1000$. As in the case of real fish, we observe that the maximal correlations corresponds to a non-zero τ_{max}, which takes a value $\tau_{max} \simeq 12$ for $T = 100$ and up to $\tau_{max} \simeq 81$ for $T = 5000$, and that grows approximately as the square root of the period, $\tau_{max}(T) \simeq T^{1/2}$.

In Figure 5.13(A) we show the direct correlations between the triangular wave velocity $v(t)$ and polarization $\phi(t)$ by means of a color density plot. As we can see, the plots are similar to those corresponding to real fish, signaling the strong

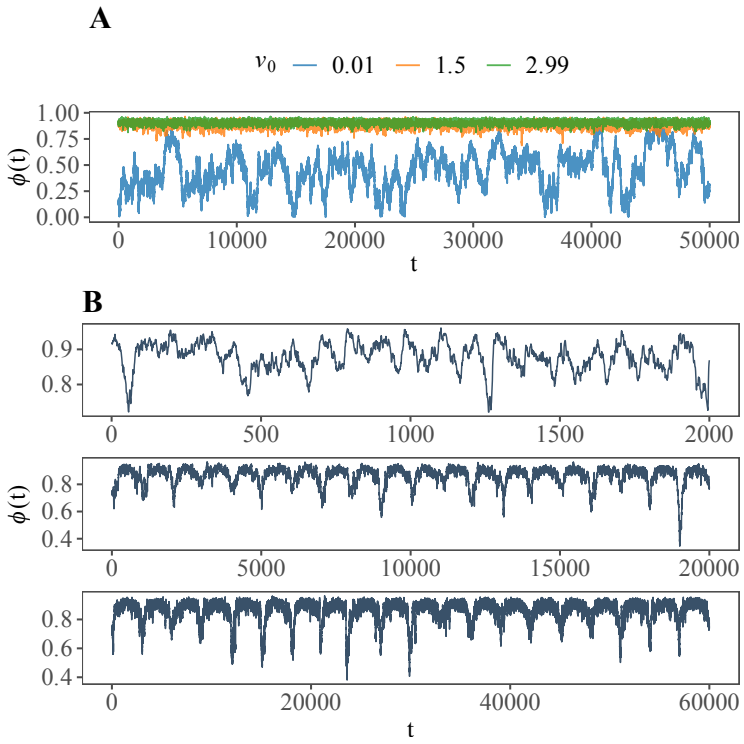


Figure 5.12: A: Instantaneous polarization $\phi(t)$ as a function of time for the Vicsek model with constant speed v_0 for different values of v_0. Noise parameter $\eta = 0.1$. **B**: Instantaneous polarization $\phi(t)$ from the Vicsek model with a time varying periodic speed given by a triangular wave of minimum value $v_{\text{min}} = 0.01$, amplitude $A = 1.49$ and different periods, $T = 100$, 500, and 1000, from top to bottom. Noise parameter $\eta = 0.1$.

correlation between these two variables, a correlation that is enhanced when the lag τ_{max} is considered, as we can observe in Figure 5.13(B).

The statistical distribution of the instantaneous polarization obtained for different values of T is shown in Fig.5.14(A), and compared with the result of our empirical observations on black neon tetra. As we can see, experimental data shows a PDF with a clear exponential increasing trend for small ϕ, followed by a very sharp decrease close to $\phi \simeq 1$. Interestingly, our model qualitatively and aptly reproduces this behavior, with higher accuracy for larger values of the period.

With consideration of the time lag velocity and polarization, we have made a standardized comparative plot, shown in Figure 5.15. The resulting shifted polarization $\phi(t + \tau_{\text{max}})$ directly follows the oscillating pattern of the velocity. In the regions of high velocity, polarization shows an almost flat behavior, compatible with the maximum, saturating polarization corresponding to large velocities, see Figure 5.13(A). The periodicity shows clearly instead in the regions of small velocity, when polarization experiences a sudden dip, which indicates that the system has briefly

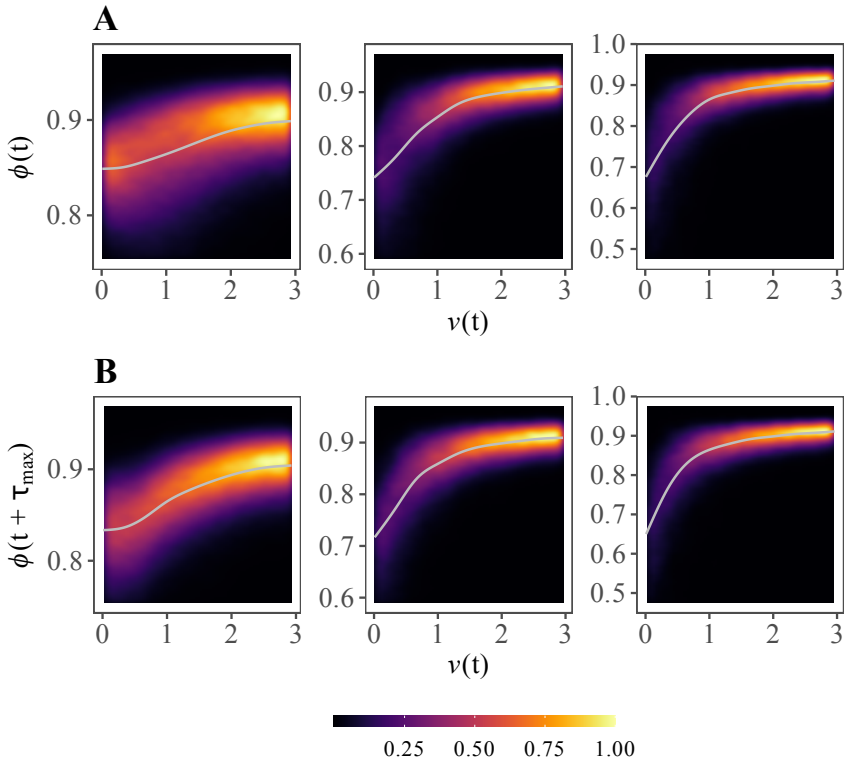


Figure 5.13: **A**: Normalized color density plot of the instantaneous polarization $\phi(t)$ versus the imposed triangular wave speed $v(t)$. Periods are, from left to right, $T = 100$, 1000, and 3000. **B**: Normalized color density plot of $\phi(t + \tau_{max})$ versus the imposed triangular wave speed.

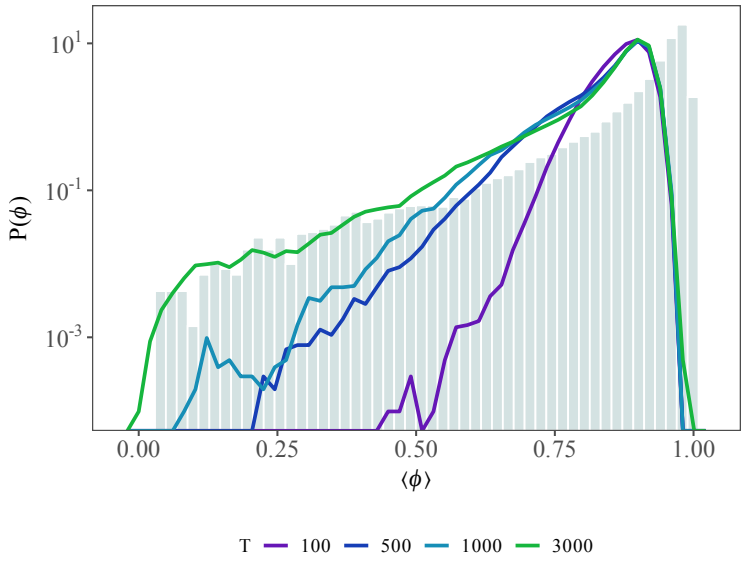


Figure 5.14: PDF of the instantaneous polarization ϕ. Light green bars refer to the empirical data, aggregated for all replicas of condition Q5N40, and continuous lines to numerical simulations of the Vicsek model with triangular wave velocity for different values of the period T.

entered the disordered state. This dynamics is reminiscent of the behavior observed in our experimental fish, Figure 5.3(B).

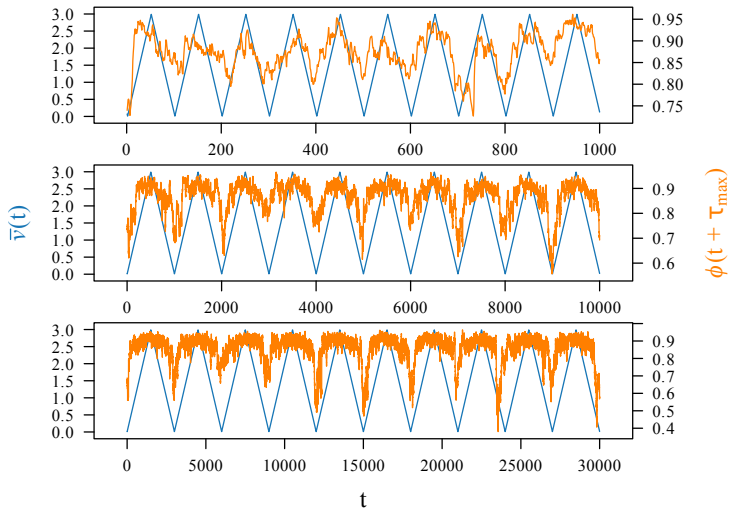


Figure 5.15: Standardized speed $v(t)$ and shifted instantaneous polarization $\phi(t+\tau_{\max})$ obtained for the Vicsek model with a triangular wave speed with different periods, from top to bottom, $T = 100$, 1000, and 3000. We show 10 oscillations cycles for each value of T.

To examine the period of the instantaneous polarization in the modified Vicsek model, as the statistics allow it, we perform in this case a Fourier analysis, computing the power spectrum of $\phi(t)$ applying the Welch method (Welch, 1967), using a time signal decomposed in $L = 64$ segments of length $M = 1500$ and overlap $D = 0.5\%$. In Figure 5.16 we show the power spectrum obtained for simulations of the Vicsek model with triangular wave velocity and different periods. As we can see, the fundamental frequency f_1, observed for each value of T, corresponding to the maximum of the power spectrum, matches perfectly with the expectation $f_1 = 1/T$, while the secondary peaks observed correspond to the harmonics $f_n = f_1/n$. That is, a periodic time varying speed induces in the Vicsek model a periodic instantaneous polarization, with the same main frequency, but with a small induced time lag which is dependent on the period.

5.3 Discussion

Here we discuss the possible effects of speed in collective motion interaction rules, and the adaptiveness of speed and polarization correlation in fish collectives. Also,

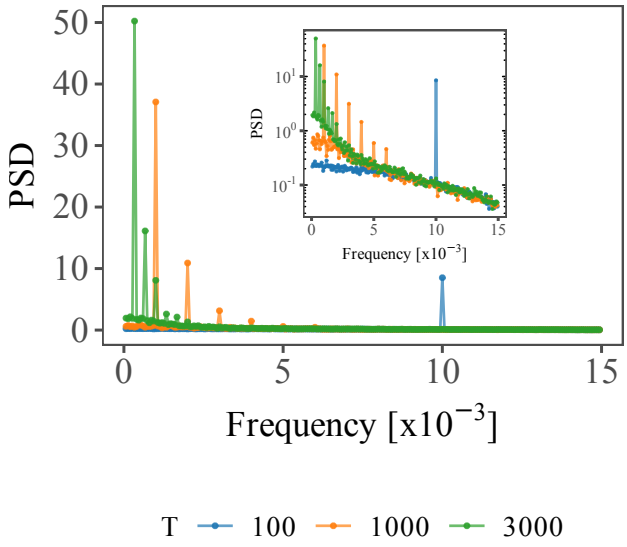


Figure 5.16: Power spectrum density for the polarization $\phi(t)$ for different triangular wave speed periods. The inset shows the same data in semi-logarithmic scale.

we address the possible impact of tank walls in fish speed modulation and compare our results with other analyses that have also looked at this potential effect. We also discuss how our modified Vicsek model with an oscillating speed can reproduce our experimental order dynamics, and what that means in respect to the simplicity of the model and the possible universality of the relationship between these parameters.

A positive correlation between speed and order has been reported in analysis of the collective motions of several fish species (Viscido et al., 2004; Gautrais et al., 2012; Calovi et al., 2018; Mishra et al., 2012; Tunstrøm et al., 2013). In both experimental work and in data-driven fish models, it has been concluded that the increase in group's speed is equivalent to increasing the alignment interaction strength, and thereby decreasing the effect of the noise in fish matching orientations. Therefore, the speed-induced phase-transition can be thought of as a competition between noise and social interactions.

From a biological perspective, speed and collective order positive correlation could be an adaptive mechanism that would ensure a faster information transfer in changing environmental conditions. This trait is part of a set of behavioral adjustments that allow a quick response to perturbations. An emblematic example is observed in predation risk avoidance: when fish feel threatened, they increase their

swimming speed and increase their ordering to collectively respond to attacks (Rieucau et al., 2015).

The analysis of the individual speed of black neon tetra shows a oscillatory, quasi-periodic behavior in time, with alternating phases of increasing and decreasing trends. Interestingly, all fish have a closely synchronized speed variation that translates in a quasi-periodic average school speed with an effective period $T \sim 2$s. The degree of order of the school, measured in terms of the global polarization, shows a similar oscillatory pattern, strongly correlated with the evolution of the average speed, and with a similar effective period. Despite this strong correlation, our experimental conditions fish show mixed polarization levels at low swimming speeds, which to the best of our knowledge, are findings not previously reported.

From an adaptive perspective, the period of fish speed oscillations could reflect the assurance that at high swimming speeds the system will always be ordered, which would translate into an efficient information transfer and a rapid group response to environmental changes. A shorter oscillation period would out-gain the system's characteristic response time and fish could exhibit lower polarization levels than predicted at high swimming speeds. This hypothesis can be supported by burst-and-coast studies that have measured the oscillation dynamics depending on their average speed, and have found that when fish are swimming at high average speeds, the burst phase is extended while the coast phase is shortened, but the periodicity of the oscillation maintains (Li et al., 2020).

To be able to disregard the influence of the tank walls on speed time series periodicity, we have studied the dependence of the velocity patterns with the distance of fish to the walls, observing a notably weak correlation, both at the level of individual fish and at the level of the centroid of the school. This precludes boundary wall effects as the cause of the oscillatory behavior, and lend support to an explanation based on the burst-and-coast swimming mechanisms.

A weak correlation between tank walls and other schooling metrics has also been reported in other fish studies. Viscido et al., 2004 compared the distance between the group centroid and the nearest tank wall, to speed, polarization and distance between individuals, and found no relationship between edge proximity and changes

in these schooling properties. Herbert-Read et al., 2011 also found weak border effects in mosquito fish accelerations and turning angles, which were instead strongly affected by the nearest neighbor position.

Although tank walls restrict group movements, its direct effect on collective motion has not been demonstrated. Border correlations with multiple schooling properties are overshadowed by the strong correlations between individuals' behavioral synchronization, a trait that facilitates groups with adaptiveness to changing environments.

We have also studied the effects of an oscillating speed on the classic Vicsek model, one of the simplest models of collective motions showing a flocking transition. The noise critical value at which phase transitions occur in the Vicsek, and even their possible nature, is dependent on combinations of speed, density and system size, which act as regulators of particles' interactions, and hence, of the randomness in the direction of motion. For example, for low densities and low particles' speeds, the phase transition is predicted to occur at low noise amplitudes, regardless of the system size (Sumpter et al., 2008; Aldana et al., 2009). On the contrary, for high densities, high speeds and large system sizes, the predicted noise critical value is higher (Grégoire and Chaté, 2004; Aldana et al., 2009). In our model, we impose a common oscillating speed, given by a triangular wave of period T. In this scenario, the global polarization closely follows the evolution of the speed, the match being better for larger periods. As observed in real fish, there is temporal lag between velocity and polarization, which can be attributed to a response time needed to integrate globally changes in speed that take place more smoothly for larger T.

Our results also coincide with those of Mishra et al., 2012, which introduced variable individual speeds to the Vicsek model. The range and the dynamics of speed changes were adapted from speed dynamics of experimental fish data where they found that when individual speed was higher, the local polar order was also higher than in areas with lower individual fish speeds. In their model, they coupled the local order with local speed to force the system to improve its local alignment when local speed increased. Here, we prove that it is not necessary to actively couple these two parameters in collective motion models, for the ordering of the system is intrinsically dependent on speed. Our results also differ from the first conclusion of

Cambui and Rosas, 2018. In their study an order-disorder transition was observed by increasing particle speed. The primary difference with our model is the introduction of an attraction zone. In their parametric space exploration, Calovi et al., 2014 concluded that increasing speed causes the effect of alignment to predominate over attraction. By forcing an attraction zone, alignment interactions could be eclipsed, and hence lose correlation with speed. Further work is needed to prove this hypothesis.

5.4 Closing remarks

Our contributions intend to provide evidence of the close relationship between swimming speed and collective ordering within fish collectives. Modulating polarization by changes in speed has an important adaptive component, as information is transferred faster and more efficiently. An intriguing question is whether this relationship is actively enforced through individuals cognitive mechanisms, or is a by-product of local interactions that favor alignment forces over others when speed is increased. We have shown that the Vicsek model can reproduce the speed and collective ordering coupling only by particles aligning with neighbors, and simulate the behavior in black neon tetra's collective motion by oscillating average speed with a sufficiently long period that allows the system to recover from its characteristic response time. This contribution furthers our understanding on the generalization of this relationship and further upholds the possibility of universality in collectives of different nature.

6 Scale-free behavioral cascades and effective leadership in schooling fish

6.1 Introduction

Sudden collective changes of state are characteristic of animal groups that collectively move (Herbert-Read et al., 2015). Sometimes they may be the result of a global alarm cue that triggers individuals reflex response. In such cases, the collective change of state is not necessarily a social response of the group. All members directly receive the stimulus and there is no actual transfer of information among them. Nevertheless, there are many examples that involve a flow of information through the collective. They could be the result of a few members of the group responding to an external cue, which are followed by the rest as information travels across the group. The most recognized example is when a nearby predator is seen by a small number of individuals that results in a sudden direction change of the entire group.

Intriguinly, it has also been shown that the collective directional switching can be also triggered spontaneously, without apparent changes in the external environment. For example, fish schools often change direction even in the absence of predators or obstacles. This poses the question as to why and how this kind of behavior arises (Attanasi et al., 2015). More so, who are the initiatiors of these spontaneous collective directional changes?

As part of answering these questions, leadership has been sometimes brought

about to rationalize the cooperative movements of animal groups by single individuals that appear to have a strong influence on the flock behavior (Couzin et al., 2005). In the case of collective travel, this translates to the timing or direction of the group's movements. In this sense, the effects of leadership have been considered in several contexts, including crowd behavior (Aubé and Shield, 2004), hierarchical leadership (Shen, 2008), linear response theory in flocking systems (Pearce and Giomi, 2016; Kyriakopoulos et al., 2016) or the emergence of complex patterns of cooperation and conflict (Smith et al., 2016).

Leadership can arise as an instinct in some animals, specially in mammals, which form a permanent hierarchical structure. In this scenario a dominant individual has won the leader role by agonistic interactions (Nagy et al., 2013). However, leadership may exhibit a switching dynamics that can even depend on contex (Nagy et al., 2010; Flack et al., 2012; Nagy et al., 2013; Chen et al., 2016). This is the case in animals that perform collective motion. There, effective leadership can come from individuals having useful information about their environment, such as the position of food or predators, not visible to the rest of the flock (Couzin et al., 2011; Ward et al., 2011). Dominance hierarchies have ben proved not to be correlated with leadership in collective decision making. For example, Nagy et al., 2013 worked with groups of 10 to 30 pigeons (*Columba livia*). They characterized their social hierarchy, expressed in dominance interactions while begin fed, and observed if they remained in the leader-follower dynamics while the group was performing free flights. They found that flight leadership is independent of the individuals' social hierarchy. By ignoring social dominance when in flight, flocks of pigeons potentially make better navigational decisions because leadership can emerge from relevant attributes, such as individual information and route fidelity (Ward et al., 2011).

Within the context of spontaneous collective directional switching, individuals that first change their direction of motion act as effective group leaders. When one or a few individuals abruptly change their dynamic, usually by modifying their velocity, the change can propagate and the group can exhibit from intermittent collective rearrangements to state transitions at the macroscopic level. The behavioral shift can occur without any perceived threat in the neighborhood, resulting in a spontaneous transition at the collective level that can be interpreted as a consequence of random

individual decisions (Couzin et al., 2005).

Spontaneous effective leadership and collective behavioral waves in groups have been studied in different animal systems (Flack et al., 2012; Ginelli et al., 2015; Attanasi et al., 2015; Rosenthal et al., 2015). Ginelli et al., 2015 focused on individual directional shifts that may trigger a collective response without an apparent external stimulus. To do so, they characterized the collective behavior of large groups of Merino sheep (*Ovis Aries*). They revealed an intermittent dynamics where sheep disperse slowly while grazing, followed by fast avalanche-like packing events, triggered by individual sheep that start running. They designed an agent-based model to explore the individual stimulus/response and found that given an individual behavioral shift, strong imitation to neighbors would create a wave of rearrangement sufficient to generate the complex collective behavior observed. High imitation could be key in individuals being able to cover large grazing areas, but rapidly regroup and protect from predation. Attanasi et al., 2015 studied spontaneous collective turns in groups of starlings *Sturnus vulgaris* and found that individuals in the peripheral tips of the flock always start the turn and display unusual deviations from the mean flock's motion over longer periods of time than other birds. The fact that the collective turn origin is so location-specific could relate to birds in the periphery experiencing higher risk for predatory attacks and less social feedback from neighbors. Rosenthal et al., 2015 reconstructed interaction networks in spontaneous startle events in golden shiners (*Notemigonus crysoleucas*), and found that interconnection among an individual's neighbors has a vital role in propagating the startle response through the group, due to the fact that multiple pathways allow for reinforcement of observations, increasing the likelihood of behavioral change. They also found that individuals near the leading and side edges of the group tend to be the most socially influential and most susceptible to social influence.

Most interestingly, these experimental studies emphasize that animal rearrangements can either spread extensively within the group or extinguish rapidly, leading to an avalanche-like type of response with a broad-tailed distribution of avalanche magnitudes (Rosenthal et al., 2015; Ginelli et al., 2015). This sort of avalanche behavior is well known in the physics literature (Fisher, 1998), where it has been discussed in magnetic materials (Zapperi et al., 1998), superconductors (Altshuler and

Johansen, 2004), plastic deformation of crystalline materials (Miguel et al., 2001), fracture phenomena (Zapperi et al., 1999), or earthquakes (Kawamura et al., 2012).

Avalanche-like collective behavior can be looked at from an adaptive perspective. For many animal collectives, such as schooling fish and flocking birds, social contagion happens very fast. The collective behavioral waves are observed in particular when individuals are under threat of attack from predators. Survival depends on how individual interactions scale to collective properties (Rosenthal et al., 2015). Both the slightest uncertainty and a slow and inefficient transfer of information are punished by decrease of cohesion, or even splitting of the group, leaving some individuals as easy prey for predators (Attanasi et al., 2015). The fact that this mechanism is observed even in the absence of external cues might reinforce its adaptive nature, as each individual does not wait to confirm the presence of an external threat, but reacts to local sudden changes.

6.1.1 Objectives

In this chapter we examine the interplay between effective leadership and behavioral cascades (avalanching behavior) by means of an empirical analysis of the movement of black neon tetra fish (*Hyphessobrycon herbertaxelrodi*), and through the theoretical analysis of a variation of the classical Vicsek model (Vicsek et al., 1995) that includes an explicit leader.

For this study we use the dataset described in Chapter 2). In our empirical analysis, we define avalanches in terms of changes in the fish heading above a given turning angle threshold, which lead to a sudden reorientation of the global trajectory of the school. We measure the statistics of avalanche initiators to explore the possible presence of consistent leadership in fish that promote substantial school rearrangements. In order to check the general effects of leadership in avalanche behavior, we consider a Vicsek-like model in which a global leader alternates a directed motion, unaffected by other individuals, with sudden variations of its direction of motion, in the spirit of run-and-tumble locomotion (Méndez et al., 2014).

6.2 Results

6.2.1 Experimental results

Quantifying avalanches in black neon tetra

Black neon tetra tend to move in a coherent and highly polarized fashion, swimming along a common and slowly changing average velocity. In this regime, the heading of the fish velocities changes almost continuously. However, at some instants of time, we can recognize swift rearrangements of the individuals' headings, that lead to a change of the average orientation of the school, accompanied by an increase of the average velocity and a decrease and a delayed increase of the global order of the school. We interpret these sudden rearrangements of individual heading as triggeres of *avalanches* of activity. In order to categorize them, we fist examine the angular variations in the heading of individual fish, defined as the *turning angle* $\varphi_i(t)$ formed by the velocity vectors $\vec{v}_i(t+1)$ and $\vec{v}_i(t)$ (see Figure 6.1A), defined in Chapter 3 and computed as

$$\varphi_i(t) = \left| \arctan \left\{ \frac{\|\vec{v}_i(t) \times \vec{v}_i(t+1)\|}{\|\vec{v}_i(t) \cdot \vec{v}_i(t+1)\|} \right\} \right|, \tag{6.1}$$

In Figure 6.1B we plot the distribution of turning angles $P(\varphi)$ for the different data replicates (hollow symbols) and for the aggregation of all three of them (red line). Extremely small values of φ are probably attributed to a low precision for very accute angles of the algorithm used in Equation 6.1. This numerical instability does not have a big effect on our analysis, as we consider much larger turning angles. Discarding these extreme values, the distributions show an extended plateau for small turning angles, corresponding to stretches of time in which fish barely change their heading and are thus compatible with movement along a smoothly winding trajectory. Instead, for values larger than 0.01 radians, the distribution starts to decrease sharply.

In order to quantitatively identify avalanches, as is customarily done in condensed matter physics (Laurson and Alava, 2006), we define a *turning threshold* φ_{th} that distinguishes small turns $\varphi < \varphi_{th}$, associated to smooth trajectories, from large turns $\varphi > \varphi_{th}$, associated to sudden rearrangements that trigger an avalanche. In

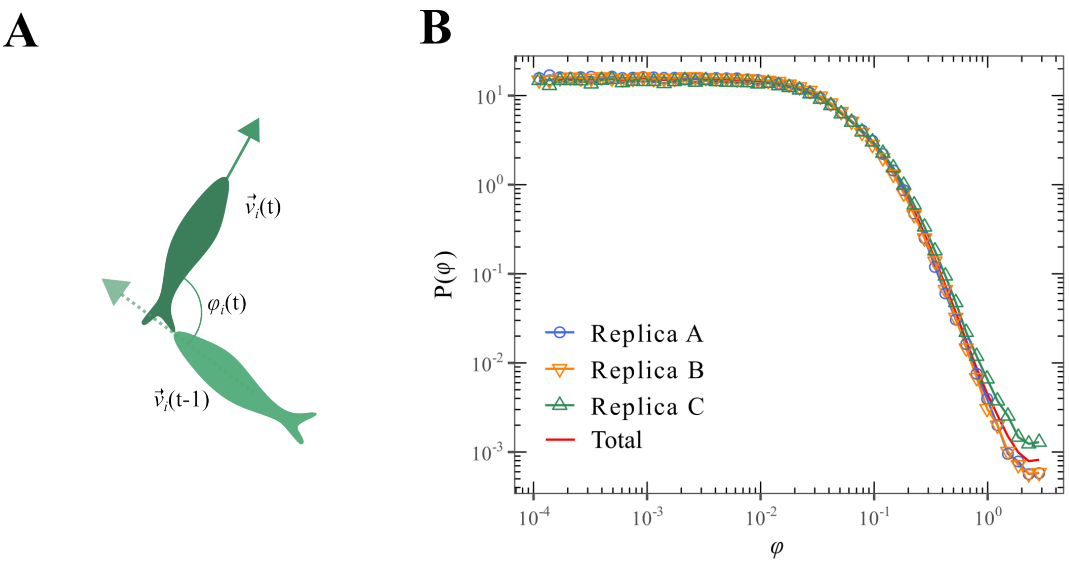


Figure 6.1: A: Example of individual turning angle φ_i from $t-1$ to t. **B**: Probability density of turning angles φ of individual fish for each replicate (hollow symbols) and for the aggregation of all three of them (red line).

Figure6.2 we plot, for a given value of the threshold, the number of *active* fish, defined as those performing a turn larger than φ_{th}, as a function of time. Here we can see the actual presence of turning avalanches, defined as trains of consecutive frames in which more than one fish is active, delimited by two frames (one at the beginning and another at the end of the train) with no active fish. These curves highlight the intermittent and heterogeneous character of avalanches, which may be rather small or can also reach relatively large sizes.

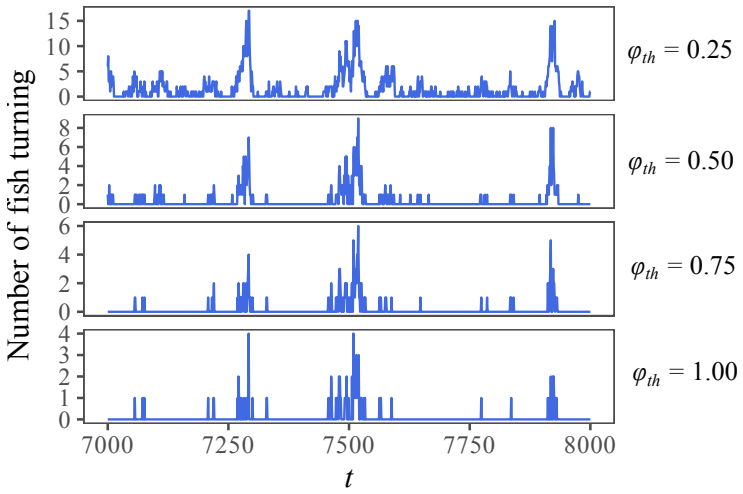


Figure 6.2: A Number of fish turning an angle larger than φ_{th} in a sequence of 1000 frames in replicate A, for different values of the turning threshold.

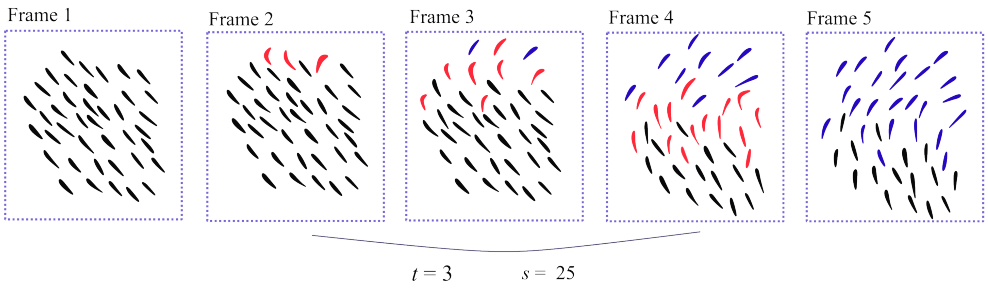


Figure 6.3: Diagram exemplifying an avalanche of duration $t = 3$ and size $s = 25$. At frame 2, three fish becom active, i.e. change their direction of motion and turn more than the turning threshold φ_{th} (red). In the following frames, some fish continue to be active while others become active. Past participants of the avalanche are colored in blue.

To study the statistics of avalanches, we compute their duration t and size s, defined as the number of consecutive time steps (frames) with at least one active fish, and the sum of the number of active fish at each time step of an avalanche, respectively. Notice that, since a fish can be active in more than one step along the duration of an avalanche, the avalanche size s is in general larger than the avalanche duration t, and can be larger than the total number of fish in the experiment. Figure 6.3 shows a diagram exemplifying an avalanche of duration $t = 3$ and size $s = 25$.

A first broad statistical characterization is given by the relative size and duration fluctuations, measured as the standard deviation σ divided by the corresponding average value μ.

Our experimental data only allows the identification of a limited number of avalanches. Indeed, in Figure 6.4(A) we plot the total number of recorded avalanches as a function of the turning threshold. From here we observe that the range of values of φ_{th} that lead to at least 1000 avalanches range approximately in the interval $[0.20, 1.50]$. In Figure6.4(B), we plot these relative fluctuations for both s and t, respectively. From this plot, we observe that relative fluctuations are only larger than 1 for threshold values within the interval between 0.1 and 1.2. We therefore restrict our analysis to the conservative threshold interval $[0.20, 1.20]$.

We consider the shape of the probability distributions of avalanche sizes, $P(s)$, and durations, $P(t)$, obtained for different values of φ_{th} (see Figures 6.5 A and B) and their correspondent cumulative probabilities distributions (see Figures 6.5 C and D),

$$P_c(s) = \sum_{s'=s}^{\infty} P(s') \qquad \text{and} \qquad P_c(t) = \sum_{t'=t}^{\infty} P(t'). \tag{6.2}$$

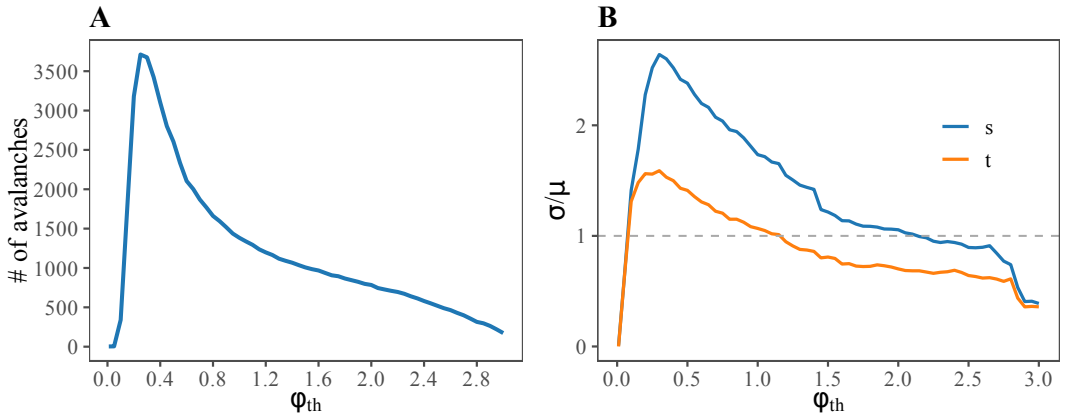


Figure 6.4: **A**: Number of avalanches observed as a function of the turning threshold. **B**: Relative fluctuations of the avalanche size (s) and duration (t) distributions as a function of the turning threshold.

From the double logarithmic scale in the plots, we can see that the distributions show long tails, compatible with a power-law behavior of the form $P(s) \sim s^{-\tau_s}$ and $P(t) \sim t^{-\tau_t}$ for small values of s and t. This power-law behavior is due to the correlated nature of turns in the fish school, which creates larger and longer avalanches than expected by random fluctuations. We corroborated this feature by comparing the results with the avalanche distributions obtained from trajectories reconstructed by randomizing the sequence of turning angles of each fish. In the latter case, one obtains a clear exponential decay, as shown in triangle symbols in Figures 6.5 C and D.

A null model of fish avalanches

We can obtain an analytical derivation of avalanche properties in the case of absence of dynamical correlations between the turning angles of fish, when the evolution of avalanches is purely determined by the independent turning probability $P(\varphi)$ of each fish. As a null model of avalanche behavior, we consider the case in which each fish independently turns a angle φ at each time step. Consider an avalanche of duration t and size s, starting at time $t' = 1$. If the avalanche lasts t time steps, it means that at least one fish turned an angle larger than φ_{th} every frame from $t' = 1$ to $t' = t$, and that no fish turned an angle larger than φ_{th} at frame $t' = t + 1$. Under these conditions, the probability that a fish turns an angle larger than φ_{th} in any

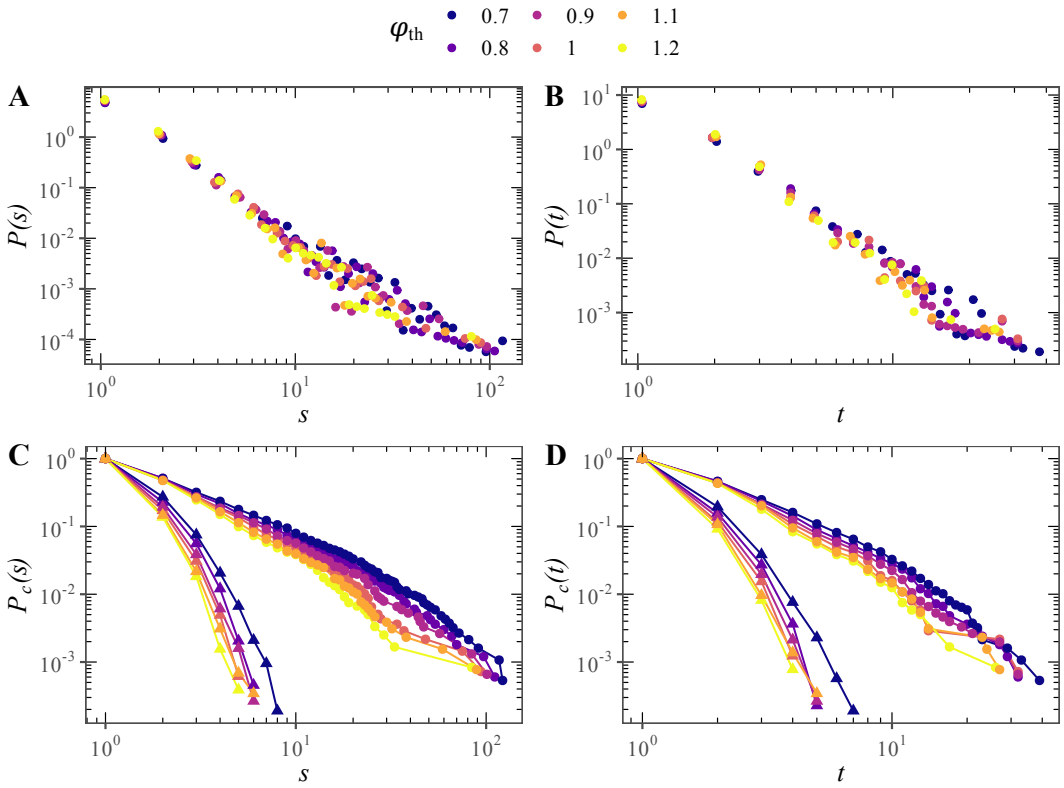


Figure 6.5: **A**: Probability distribution of the avalanche sizes $P(s)$ for different values of the turning threshold φ_{th}. **B**: Probability distribution of the avalanche durations $P(t)$ for different values of the turning threshold φ_{th}. **C**: Cumulative probability distribution of the avalanche sizes $P_c(s)$ for different values of the turning threshold φ_{th}. **D**: Cumulative probability distribution of the avalanche durations $P_c(t)$ for different values of the turning threshold φ_{th}. In both **C** and **D** points represent the actual empirical distributions, while triangles correspond to distributions obtained by randomizing the turning angles in the trajectory of each fish.

frame is

$$q = \int_{\varphi_{th}}^{\pi} P(\varphi)\, d\varphi, \tag{6.3}$$

and the probability that at least one fish turns an angle larger than φ_{th} in a given frame is

$$Q = 1 - (1 - q)^N, \tag{6.4}$$

where N is the number of fish. Therefore, the normalized probability that an avalanche lasts for t frames in this null model is

$$P_0(t) = \frac{(1-Q)Q^t}{\sum_{t'=1}^{\infty}(1-Q)Q^{t'}} = (1-Q)Q^{t-1}, \qquad t = 1, 2\ldots, \infty, \tag{6.5}$$

where we consider that avalanches have a minimal duration of one frame. That

is, in the uncorrelated null model, the avalanche duration distribution has an exponential form, with average avalanche duration $t_0 = \sum_{t=1}^{\infty} tP_0(t) = 1/(1-Q)$.

Consider now a frame in an avalanche of finite duration. In this frame, at least one fish turned an angle larger than φ_{th}, therefore the probability of observing the s_1 large turns in this frame is

$$p_1(s_1) = \frac{1}{Q}\binom{N}{s_1} q^{s_1}(1-q)^{N-s_1}, \qquad s_1 = 1, 2, \ldots, N, \tag{6.6}$$

If the avalanche has duration t, at each frame a number s_1 of fish, distributed with the probability in Eq.(6.6), will turn a large angle. Therefore, the distribution of sizes in avalanches of duration t, $P_0(s|t)$ will be given by the convolution of the probability in Eq.(6.6) t times with itself. The form of this expression is hard to compute. However, we can approximate the avalanche size distribution as follows: Since Eq.(6.6) is similar to a binomial distribution, it is bell-shaped and centered at the average value

$$\bar{s}_1 = \sum_{s_1=1}^{N} s_1 p_1(s_1) = \frac{Nq}{Q}. \tag{6.7}$$

Therefore, the average size of an avalanche of duration t is

$$\bar{s}_t = \frac{Nq}{Q}t, \tag{6.8}$$

linear with t. Assuming that the relation between size s and duration t is tight, given the bounded distribution $p_1(s_1)$, we can use relation $s \simeq \frac{Nq}{Q}t$ and the distribution $P_0(t)$ from Eq.(6.5) to obtain the probability transformation $P_0(t)dt = P_0(s)ds$, leading to

$$P_0(s) \simeq \frac{1-Q}{Nq} e^{-sQ\ln(1/Q)/(Nq)}, \tag{6.9}$$

that is, an exponential decay with a characteristic size

$$s_c = \frac{Nq}{Q\ln(1/Q)}. \tag{6.10}$$

as we observe in Figures 6.5C and D.

Scaling behavior of avalanche sizes and durations

Upon closer scrutiny of avalanche sizes and durations probability distributions in Figure6.5C and D, we can also observe that, for sufficiently large φ_{th}, the initial power-law behavior of the size distributions is followed by a faster decay for s larger than a characteristic size s_c that appears to be a decreasing function of the threshold φ_{th}.

Inspired by the observations in other avalanche systems (Laurson et al., 2009) and in models of self-organized criticality (Pruessner, 2012), we can assume that, for different values of the threshold, the size distributions exhibit a scaling behavior of the form

$$P(s) = s^{-\tau_s} G_s\left(\frac{s}{s_c(\varphi_{th})}\right) \tag{6.11}$$

where the scaling function $G_s(z)$ is constant for small $z \ll 1$ and decays rapidly to zero for $z \gg 1$. In analogy with avalanches in condensed matter and critical phenomena (Yeomans, 1992; Cardy, 1996) we make the ansatz for the behavior of the size cut-off $s_c(\varphi_{th}) \sim \varphi_{th}^{-\sigma_s}$, where σ_s is a characteristic exponent. We can estimate the values of the exponents by noticing that Eq.(6.11) implies, for the cumulative distribution $P_c(s) = s^{-\tau_s+1} F_s\left(s\varphi_{th}^{\sigma_s}\right)$, where $F_s(z)$ is another scaling function. The previous expression can be rewritten as

$$\varphi_{th}^{\sigma_s(1-\tau_s)} P_c(s) = F_s'\left(s\varphi_{th}^{\sigma_s}\right), \tag{6.12}$$

where $F_s'(z) = z^{-\tau_s+1} F_s(z)$. Eq.(6.12) implies that, when plotting the rescaled distribution $\varphi_{th}^{\sigma_s(1-\tau_s)} P_c(s)$ as a function of the rescaled size $s\varphi_{th}^{\sigma_s}$, with the correct exponents τ_s and σ_s, plots for different values of φ_{th} should collapse onto the same universal function $F_s'(z)$.

Using this idea, one can estimate numerically the exponents τ_s and σ_s as those that provide the best collapse of the data rescaled using Eq.(6.12) for the different values of φ_{th}. Data collapse is performed by considering the difference of the curves for the different values of φ_{th} and choosing the exponents τ_s and σ_s that minimize

this difference. To compute the difference, we locate the interval of values of $s\varphi_{th}^{\sigma_s}$ common for all φ_{th}. In this interval, we compute a spline of order k for each quantity $\varphi_{th}^{\sigma_s(1-\tau_s)}P_c(s)$ and interpolate a fixed number n of equispaced points. The difference is defined as the sum of the variances of the values of $\varphi_{th}^{\sigma_s(1-\tau_s)}P_c(s)$ in each point of the interpolation, for the different values of φ_{th}. In the results presented here, we consider splines of order $k = 2$ and interpolate $n = 10$ points for each $P_c(s)$ curve.

Following this approach, using values of φ_{th} in the interval $[0.7, 1.2]$, we estimate the exponents $\tau_s \simeq 2.03$ and $\sigma_s \simeq 3.03$. In Figure6.6A we show the data collapse for Eq.(6.12) obtained for the cumulated size distributions using these values. Different intervals of the turning threshold provide slightly different values of the exponents, from which we estimate the average exponents quoted in Table 6.1. The same procedure can be applied to the duration distribution, see Figure6.6B, where now the cumulative duration distribution $P_c(t)$ fulfills Eq.(6.12) with the corresponding exponents τ_t and σ_t. In the same interval of thresholds we find $\tau_t \simeq 2.33$ and $\sigma_t \simeq 1.59$, see Figure6.6B, while the average exponents are given again in Table 6.1.

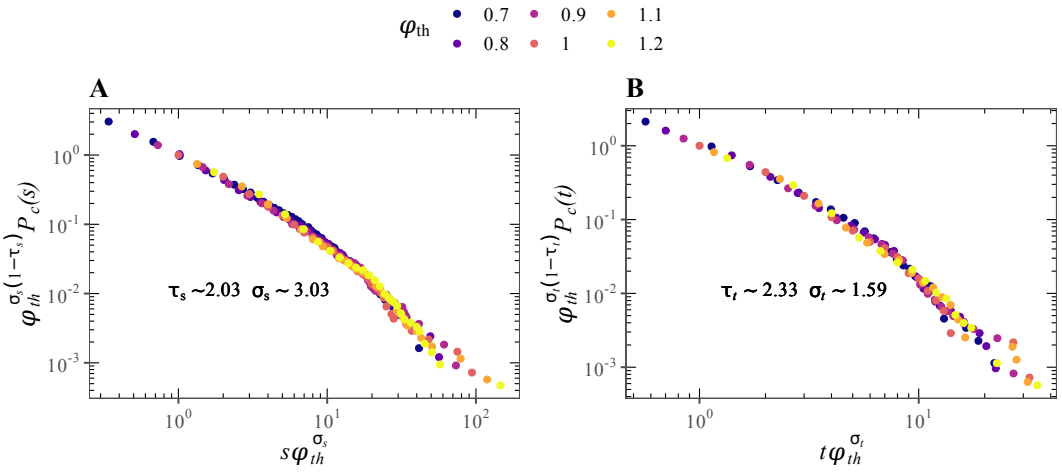


Figure 6.6: **A**: Check of the scaling of the cumulative size distribution with the turning threshold, as given by Eq.(6.12). **B** Check of the scaling of the cumulated time distribution with the turning threshold, as given by Eq.(6.12.)

τ_s	σ_s	τ_t	σ_t
2.0(1)	3.1(3)	2.4(1)	1.70(4)

Table 6.1: Summary of scaling exponents for the avalanche size and duration distributions obtained from observations of black neon tetra

We can check the validity of these results considering that, for small values of s and t, the distributions $P(s) \sim s^{-\tau_s}$ and $P(t) \sim t^{-\tau_t}$ imply that the average size of avalanches of duration t, $\bar{s}_t$, takes the form

$$\bar{s}_t \sim t^m, \qquad \text{with} \qquad m = \frac{\tau_t - 1}{\tau_s - 1}. \tag{6.13}$$

In Figure6.7 we represent the empirical average avalanche size $\bar{s}_t$ as a function of the duration t. For the different values of the turning threshold considered, we estimate numerically that $\bar{s} \sim t^{1.31}$. This observation is in good agreement with the expression in Eq.(6.13), which, using the values from Table 6.1, yields $m = 1.4(2)$. In Figure6.7(A) we also show the average avalanche size observed in randomized avalanches, which shows a linear dependence as expected theoretically, see Eq.(6.8). This last result highlights the relevant effect of turning angle correlations in real fish.

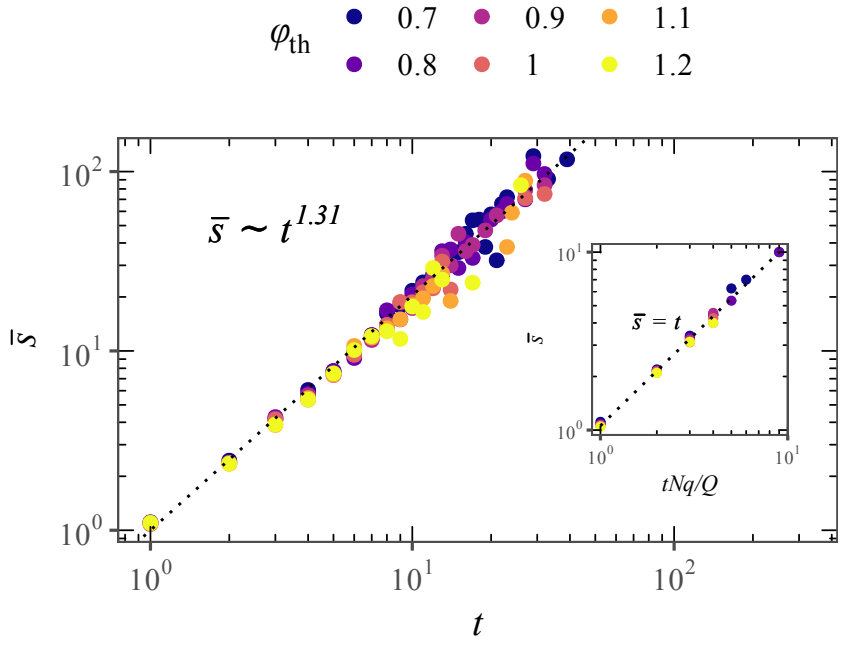


Figure 6.7: A: Average size $\bar{s}_t$ of avalanches of fixed duration t as a function of t. The main plot shows the empirical data. The inset presents the results from a randomization of the turning angles in each fish trajectory. In this case, we plot the average duration as a function of the theoretical prediction Nqt/Q, see Eq. (6.8).

Effective leadership and avalanche behavior

In order to explain the origin of the avalanche behavior observed in our empirical data, we consider the possibility that avalanches are triggered by some initiator or effective *leader*, which consistently starts the large turning rearrangements that lead to the formation of an avalanche. While several definitions of leadership have been proposed within the field of collective animal motion (Strandburg-Peshkin et al., 2018), here we use a measure explicitly devised to detect the presence of preferential initiators of avalanches. We consider the originator of an avalanche as the fish that performs the first large heading turn in the evolution of the avalanche. As more than one fish can be active in any frame, we consider as initiators all active fish in the first frame of an avalanche. We define the *leadership probability* χ_i of fish i in a given data replicate as the ratio of the number of avalanches in which the fish i is active in the first frame, divided by the total number of avalanches in which fish i participates. An example of a fish that participates in two avalanches, and initiates one of those, is shown in Figure 6.8.

The calculation is restricted to sufficiently large avalanches, of duration larger than 5 frames. In Figure 6.9 we plot the value of χ_i computed for each one of the $N = 40$ fish in each series, for different values of the turning threshold φ_{th}. As we can see, the leadership probability shows an important variation among fish. Moreover, for the largest values of φ_{th} considered, the leadership probability can take values up to 0.60, indicating that some fish initiate more than half of the avalanches in which they participate.

In order to quantify the relevance of the values of χ_i obtained, and ascertain that they are not the effect of random fluctuations in the activity of the fish, given our small populations, we compare our empirical estimates with the results obtained in a null model in which the turns performed by fish are completely independent.

In the avalanche null model defined in Section 6.2.1, consider a fish that participates in a given avalanche. To estimate its leadership probability we have to compute the probability that it leads the avalanche (i.e. it is active in its first time step), provided that it participates in it. To compute it, we use Bayes' theorem to write

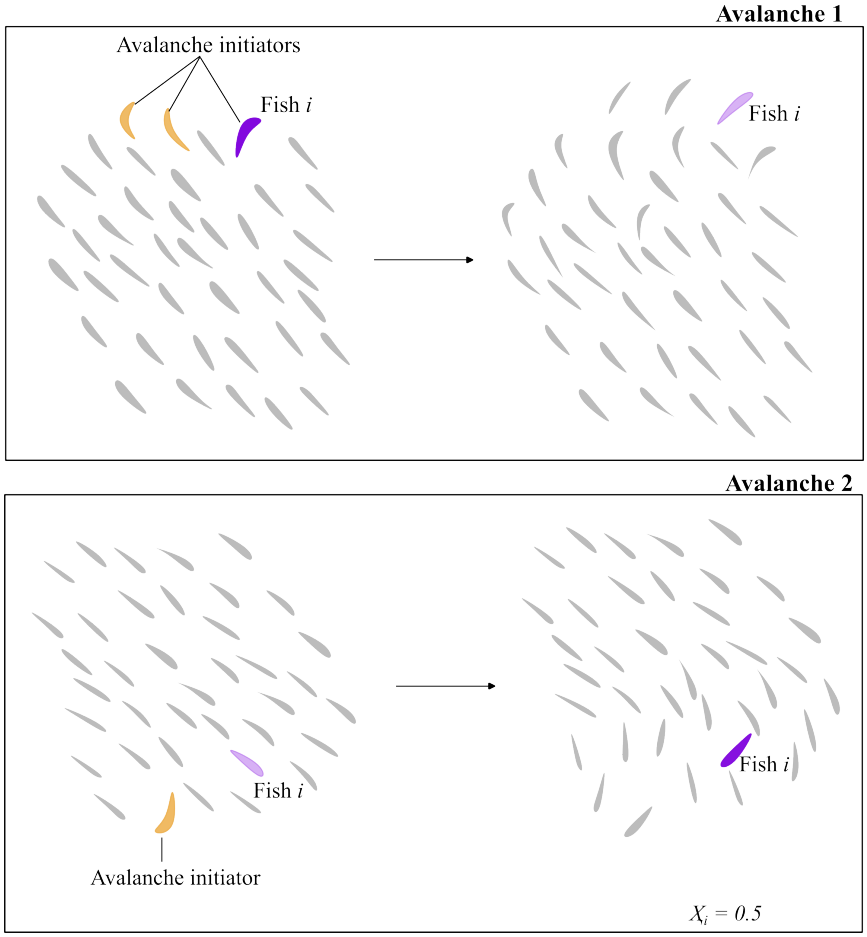


Figure 6.8: Leadership Probability example diagram. Fish i, along with two other fish, originates Avalanche 1. In Avalanche 2, fish i participates in the avalanche but it does not initiate it. In this example of two avalanches, fish i leadership probability is $\chi_i = \frac{1}{2} = 0.5$.

$$P(p)P(l|p) = P(l)P(p|l), \tag{6.14}$$

where $P(p)$ is the probability that the fish participates in a given avalanche, $P(l|p)$ the probability of leading an avalanche in which it participates (the probability we are seeking), $P(l)$ the probability of leading an avalanche, and $P(p|l)$ the probability that a fish participates in an avalanche provided that it leads it. Obviously, $P(p|l) = 1$. To estimate the rest of probabilities, we need information about the duration t of the avalanche. Thus, we have $P(p) = 1 - (1 - q)^t$, the probability that the fish turns at least once in the development of the avalanche, and $P(l) = q$, the probability that the fish is active (performs a large turn) in the first time step of the avalanche. Therefore, from Eq.(6.14) we obtain

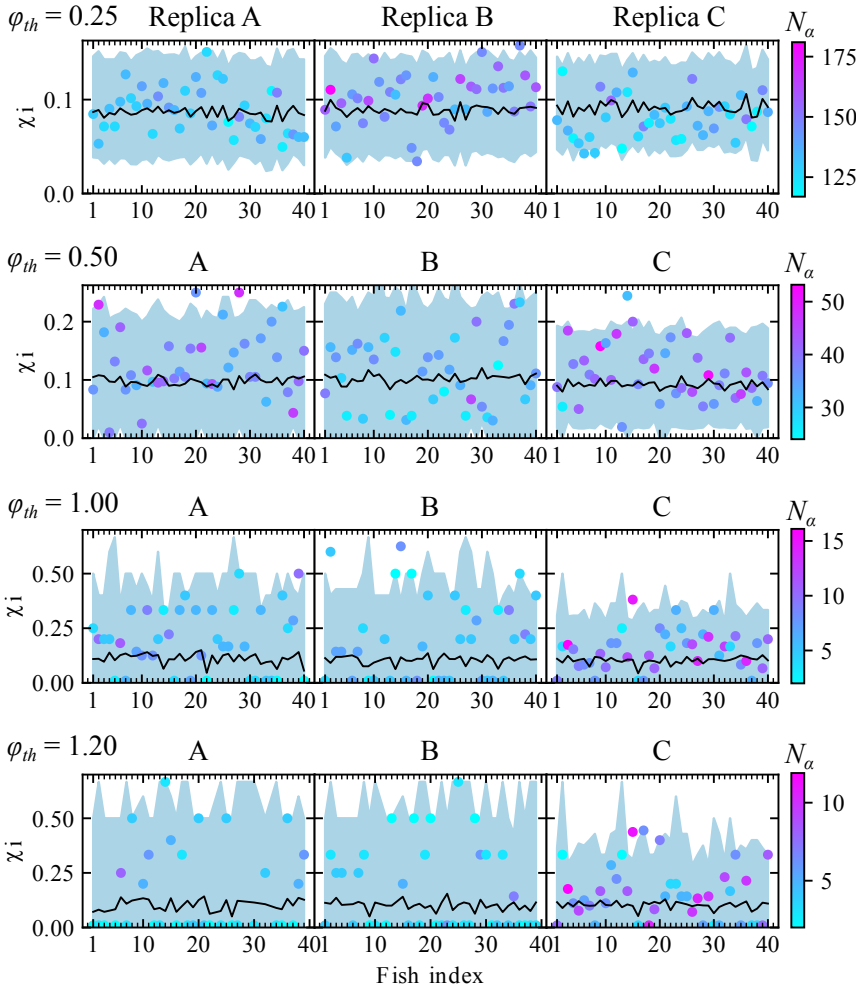


Figure 6.9: Leadership probability χ_i computed for each fish i in the three data replicates considered A, B and C (rows, left to right) and for different values of the turning threshold $\varphi_{th} = 0.25, 0.5, 1.0$ and 1.2 (columns, top to bottom). Symbols are color-coded with the number N_α of actual avalanches in which each fish participates. Full lines represent the average leadership probability in a null model of uncorrelated avalanches. The shadowed regions represents the 99% confidence interval of this value.

$$P(l|p) = \frac{P(l)P(p|l)}{P(p)} = \frac{q}{1-(1-q)^t}. \tag{6.15}$$

Within this null model, consider a fish that participates in N_a avalanches, each of duration t_α, $\alpha = 1, \ldots, N_a$. The probability of leading any of these avalanches is $p_\alpha = q/[1-(1-q)^{t_\alpha}]$. Therefore, the probability $P(\ell)$ of leading ℓ of the N_a avalanches is given by a Poisson binomial distribution, representing the probability distribution of a sum of independent Bernoulli trials that have different success probabilities p_α (Hoeffding, 1956). The Poisson binomial distribution has a rather convoluted form, but its mean and variance can be easily expressed as (Hoeffding, 1956)

$$\mu = \sum_\ell \ell P(\ell) = \sum_\alpha p_\alpha, \qquad \sigma^2 = \sum_\ell \ell^2 P(\ell) - \left[\sum_\ell \ell P(\ell)\right]^2 = \sum_\alpha p_\alpha(1-p_\alpha). \tag{6.16}$$

The average leadership probability of a fish in this null model is thus given by $\chi_0 = \ell/N_a$, where $\ell = \sum_\alpha p_\alpha$ is the average number of avalanches led by the fish.

In Figure 6.9 we show the actual values of χ_i computed for each fish. The full line and shaded region represents the null-model average value χ_i^0 computed for each fish, taking into account the number of avalanches in which it participates, and its 99% confidence interval, respectively.

Our results indicate that, with the exception of replicate C, in all replicates and for all values of the turning threshold considered, several fish have an unusually very large probability to initiate an avalanche, much larger than the value expected from pure random fluctuations. We can associate them to effective leaders of the school, which initiate with large probability the avalanches in which they participate.

More information can be obtained by considering the evolution of the leadership probability as a function of the turning threshold for each fish in each time series, as seen in Figure 6.10. From this plot we can confirm, first of all, that some fish never initiate an avalanche ($\chi_i = 0$) for large values of the turning threshold, while others consistently start much less avalanches than they should by mere random fluctuations. Some other fish behave as initiators for some range of values of the

turning threshold. Finally, some fish reliably initiate a large number of avalanches, much more than they should by pure randomness. These fish can be identified as consistent effective leaders, which trigger a large majority of the avalanches, independently of the value of the threshold used to quantitatively define them.

6.2.2 Modeling avalanches in the presence of leaders

Model definition

To explore the effects of the presence of leadership in the avalanche behavior of schooling fish, we consider as the simplest modeling scenario a variation of the classic Vicsek model (Vicsek et al., 1995) in which we introduce an effective leader.

As described in Chapter 4, the Vicsek model (Vicsek et al., 1995; Ginelli, 2016) consists on particles that interact among them by trying to align their instantaneous velocity with the average velocity of the set of nearest neighbors inside a circular region of radius R. A noise source of strength η, representing physical or cognitive difficulties in gathering or processing local information, allows the formation of an ordered (*flocking*) phase at low noise intensities, and of disordered (*swarming*) states at high enough noise values.

In the variation of the Vicsek model we consider, one of the SPPs, say particle 1, plays the role of a leader which influences the orientation of the rest of particles in the system, independently of their distance to it. The velocity of the leader, $\vec{v}_1(t) = \vec{v}_L$ is not affected by the behavior of its neighbors, and therefore its heading remains constant $\theta_1(t) = \theta_L$ over time, which represents a privileged direction it wants to follow.

In the heading update rule in Eq. 4.2 of the classic Vicsek model, the average velocity of the neighbors, $\vec{V}_i(t)$, is replaced by the average $\vec{V}_i^L(t)$ as follows

$$\theta_i(t + \Delta t) = \Theta\left[\vec{V}_i^L(t)\right] + \eta\ \xi_i(t). \tag{6.17}$$

In the classical Vicsek model, $\vec{V}_i(t)$ is calculated considering the particles located inside a circular area $\mathcal{V}_i$ of radius R centered at particle i position. Here, $\vec{V}_i^L(t)$ is

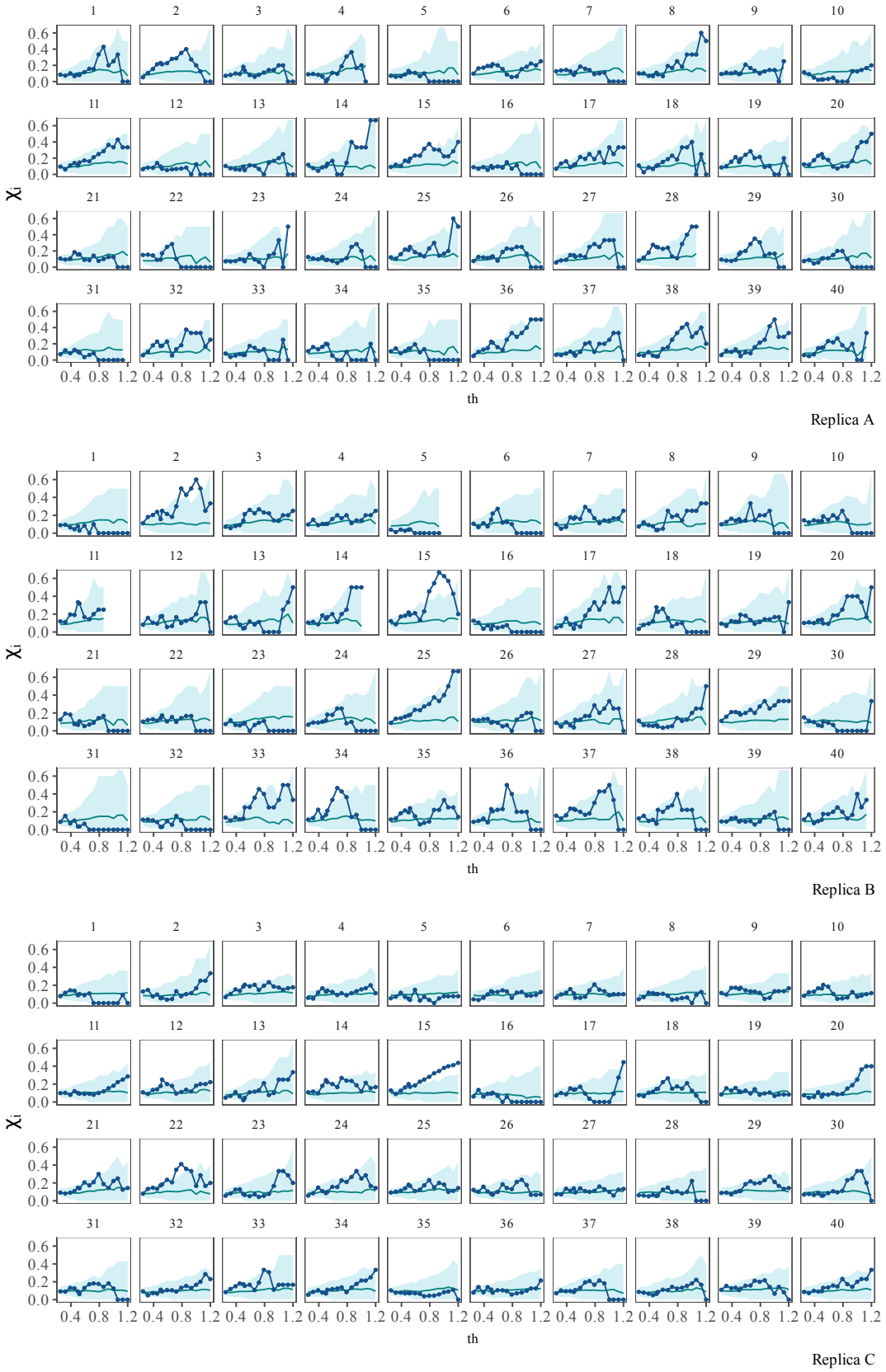


Figure 6.10: Individual fish leadership probability χ_i as a function of φ_{th} in the three replicates A, B and C. Symbols are color-coded with the number N_α of actual avalanches in which each fish participates. Full lines represent the average leadership probability in a null model of uncorrelated avalanches. The shadowed regions represents the 99% confidence interval of this value.

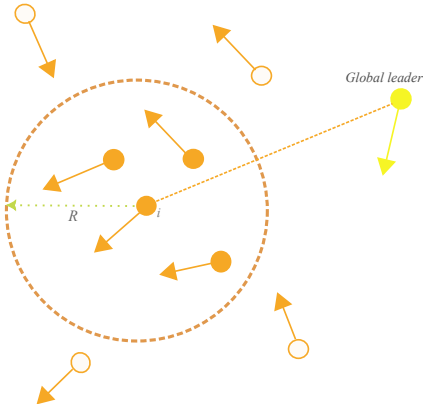


Figure 6.11: Schematic of the modified Vicsek model with a global leader that influences the orientation of particle i, independently of the distance to it.

computed in the set $\mathcal{V}_i^L = \mathcal{V}_i \cup \{1\}$, including the global leader and all the particles in the local neighborhood of i (see Figure 6.11).

Simulations of the model are performed in square boxes of different size L with periodic boundary conditions. We fix the density of particles $\rho = N/L^2 = 1$, the radius of interaction $R = 1$, and the constant speed of the SPPs $v_0 = 0.03$.

The presence of an unperturbed leader has the effect of suppressing the disordered phase exhibited by the classic Vicsek model. As we can see in Figure 6.12, while for the classical model the transition becomes sharper when increasing the systems size L, the leader induces an ordered state for any value of η, with an order parameter fairly independent of system size and vanishing only in the limit of maximum noise $\eta = 1$.

Avalanche behavior in response of leader perturbations

In this section, we focus our attention on the system-wide perturbations that are induced by changes in the preferred direction of motion of the leader. To analyze them, we consider a random reorientation of the leader's heading by an angle $\Delta\theta_L$ (see Figure 6.13), performed in the steady state corresponding to a given value of the noise intensity η, and measure the subsequent rearrangements that this perturbation induces in the heading of the rest of fish, as given by the turning angle $\varphi_i(t) = \theta_i(t+1) - \theta_i(t)$ projected on the interval $[0, \pi]$.

In Figure 6.14 we represent the probability density of SPPs turning angles $P(\varphi)$

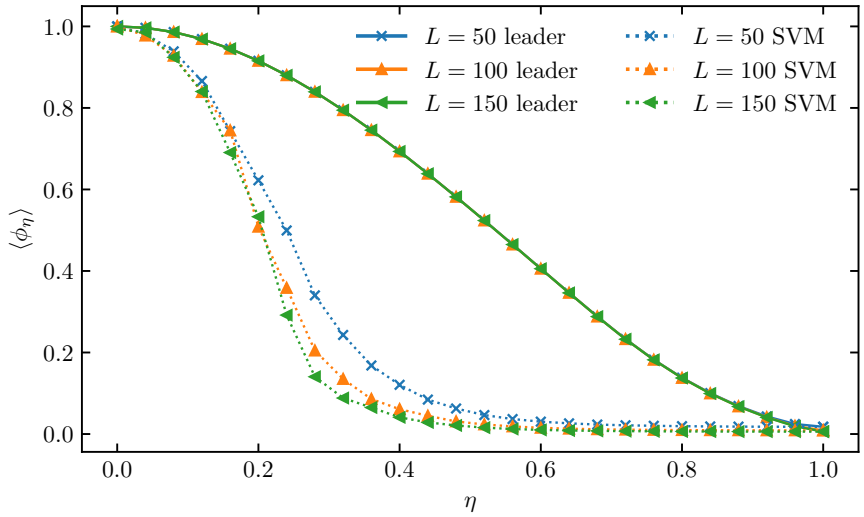


Figure 6.12: Average order parameter $\phi(\eta)$ as a function of noise intensity in the classic standard Vicsek model (SVM) and the Vicsek model with a non-rotating global leader for different system sizes.

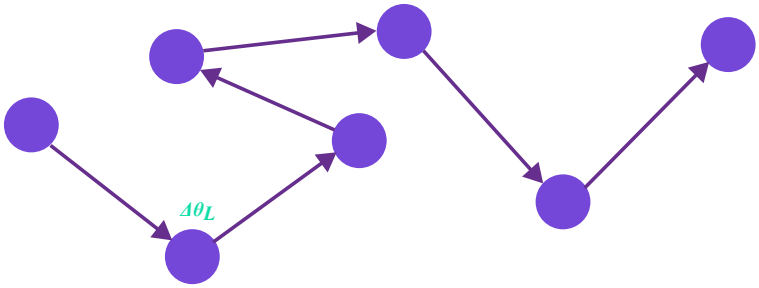


Figure 6.13: Schematic of run-and-tumble of turning leader in the modified Vicsek Model

in the steady state, for different values of the noise intensity η. In this plot we consider the model with a fixed, non-turning leader (dashed lines), and the case of a periodically perturbed leader (full lines), in which the leader experiences a random rotation $\Delta\theta_L$ of its heading, uniformly distributed in the interval $[-\pi, \pi]$, every 250 time steps, a time lapse larger than the maximum avalanche duration recorded in simulations.

As we can see, for fixed η, the two distributions are almost identical for small φ, while they differ drastically regarding the behavior of the tails beyond a given cut-off turning angle $\varphi_c(\eta)$. A numerical analysis performed for different values of L allows to estimate this cut-off as $\varphi_c(\eta) \simeq 2.4\pi\eta$. The presence of this turning angle cut-off, not available in empirical data, permits to distinguish the changes of heading due to the effect of the leader perturbations, and suggests that the proper definition of avalanches should consider turning thresholds larger than the cut-off $\varphi_c(\eta)$. In

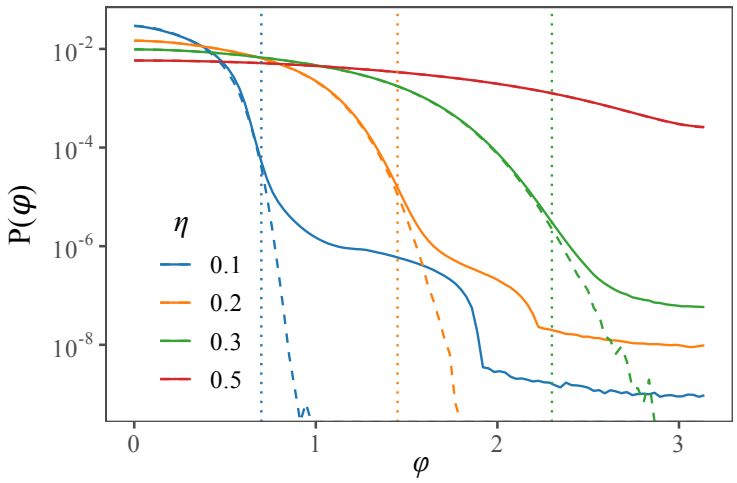


Figure 6.14: Probability density of the turning angles φ for the Vicsek model with a leader, for different values of the noise intensity η, in a system of size $L = 220$. Dashed lines correspond to a non-turning leader. Full lines represent a leader perturbed periodically every 250 time steps. Vertical lines, indicating the departure of the distributions for perturbed and non-perturbed leaders, are estimated at a value $\varphi_c(\eta) \simeq 2.4\pi\eta$.

the following, we will fix the value of the threshold to $\varphi_{th}(\eta) = 2.5\pi\eta$. We notice that for large $\eta \geq 0.5$, the angular distributions with and without perturbations are identical, compatible with a large noise masking external perturbations and making avalanches non discernible.

In Figure6.15 we plot a sample of the number of SPPs that turn an angle larger than $\varphi_{th}(\eta)$ as a function of time. This curve emphasizes the heterogeneous character of the avalanche sizes in response to the leader's changes of direction, akin to what is observed in fish schools: Sometimes a perturbation is followed by a small number of SPPs reorientations; but other times, it triggers the reorientation of a large number of particles. As expected, the strength of the effects of the leader perturbations decreases with increasing noise, indicating that interesting avalanche behavior will only occur for moderate levels of noise.

We compute the cumulative probability distributions $P_c(s)$ and $P_c(t)$, defined in Equation (6.2), of observing an avalanche of size and duration larger than s and t, respectively, plotted in Figure6.16(A) and (B) for a turning threshold $\varphi_{th}(\eta)$ and different values of η. As we can see from these plots, the values $\eta = 0.2$ and 0.3 lead to size and duration distributions analogous to that observed in rearrangement avalanches in black neon tetra schools, with a shape that can be approximated by a

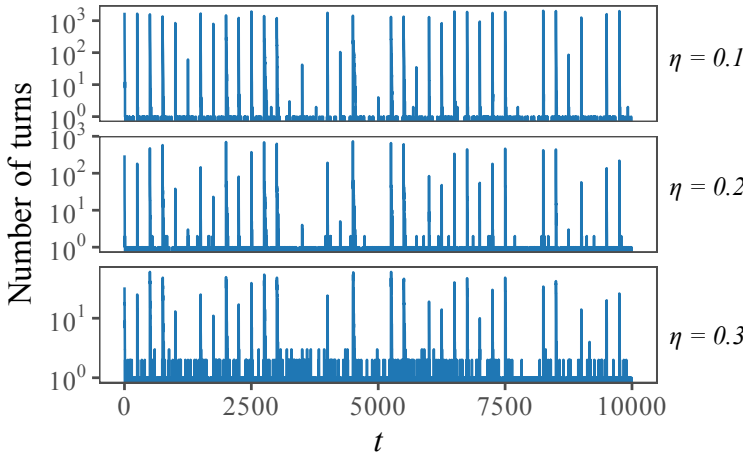


Figure 6.15: Number of SPPs turning an angle larger than $\varphi_{th}(\eta) = 2.5\pi\eta$ in a sequence of 10000 simulation time steps in the Vicsek model with a periodically perturbed leader, for different values of η.

power-law form for intermediate values, followed by a crossover to a sharp decrease for large s and t above a characteristic size or duration. The behavior for $\eta = 0.1$ is more complex, probably due to the fact that for small noise one expects a fairly homogeneous response with many SPPs following a leader perturbation. We thus discard this value in the following analysis.

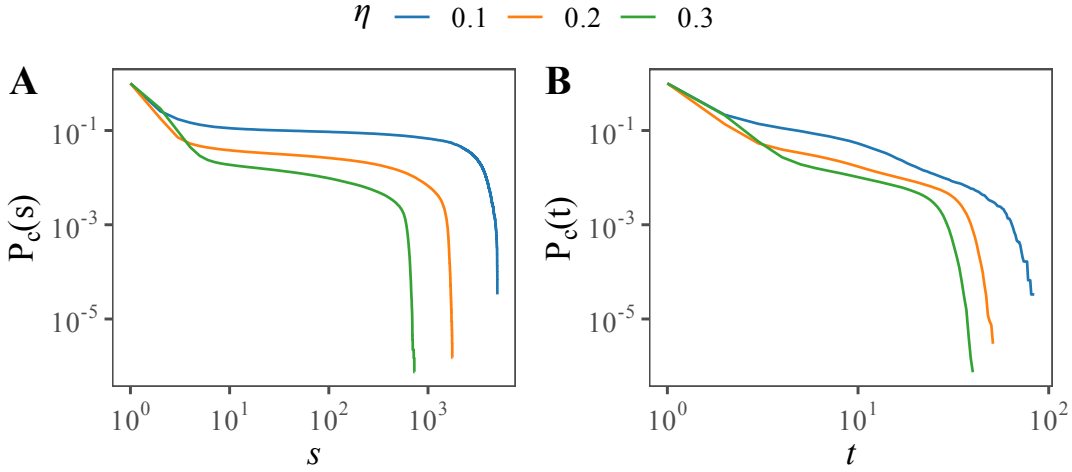


Figure 6.16: **A**: Cumulated distribution of sizes $P_c(s)$ of avalanches induced by a periodically perturbed leader in a system of size $L = 220$ with turning threshold $\varphi_{th}(\eta)$, for different values of the noise intensity. **B**: Cumulated distribution of durations $P_c(t)$ of avalanches induced by a periodically perturbed leader for different values of the noise intensity.

The fact that we work now with a numerical model, allows us to explore the behavior of the system for different systems sizes L at a fixed turning threshold,

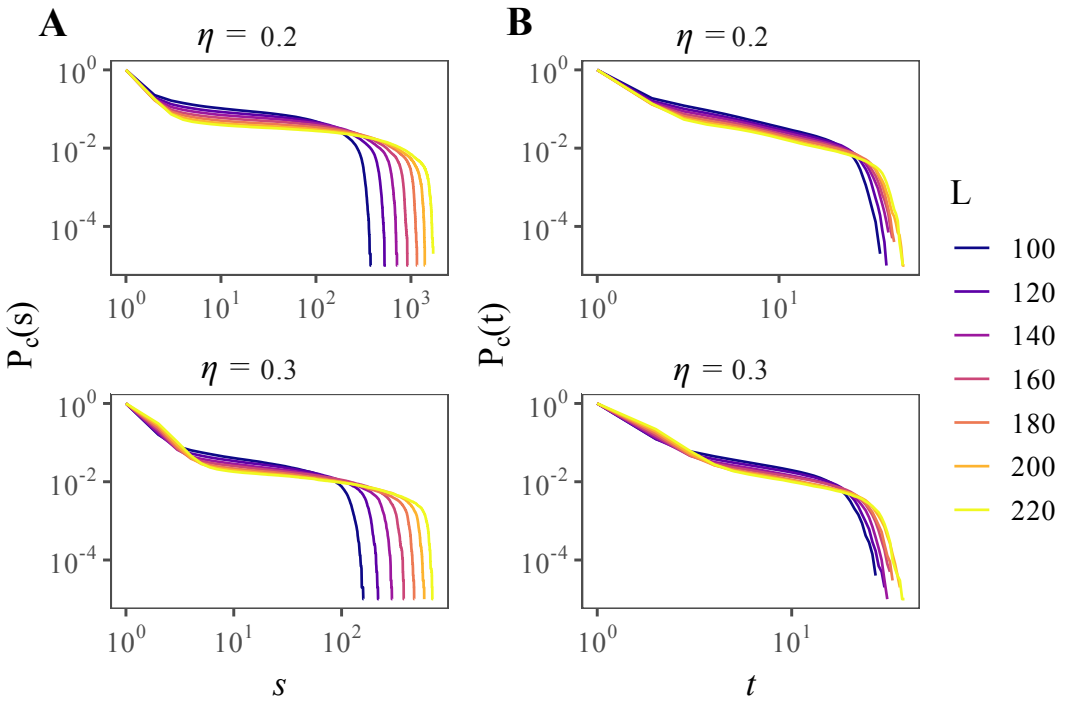


Figure 6.17: **A**: Cumulative probability distribution of sizes $P_c(s)$ and durations $P_c(t)$ of avalanches induced by a perturbed leader in a system with $\eta = 0.2$, turning threshold $\varphi_{th}(\eta)$ and different values of L. **B**:Cumulative probability distribution of sizes $P_c(s)$ and durations $P_c(t)$ of avalanches induced by a perturbed leader in a system with $\eta = 0.3$, turning threshold $\varphi_{th}(\eta)$ and different values of L.

which was not possible in our fixed size empirical data. In Figure6.17(A) we plot the cumulative size and duration distributions in avalanches in the Vicsek model with a turning leader for a turning threshold $\varphi_{th}(\eta)$, $\eta = 0.2$ and $\eta = 0.3$ and different system sizes. As we can observe, the behavior of the distributions is analogous to that observed in real fish schools, compatible with a power-law decay but that are now truncated by a size and time cut-offs that are functions of the system size L.

Inspired again by self-organized criticality (Pruessner, 2012), we can assume now that the distributions obey a finite-size scaling form

$$P(s) = s^{-\tau_s} G_s\left(\frac{s}{L^D}\right), \quad P(t) = t^{-\tau_t} G_t\left(\frac{t}{L^z}\right). \tag{6.18}$$

where D and z are new characteristic exponents that define the characteristic size $s_c(L) \sim L^D$ and time $t_c(L) \sim L^z$ as a function of the system size (Pruessner, 2012; Cardy, 1988).

The better statistics in numerical simulations allow to estimate the characteristic exponents applying the more precise moments analysis technique (De Menech et al.,

1998), defined as follows.

The finite-size scaling (FSS) method (Cardy, 1988) assumes that the dependence on system size L of the avalanche size and time distributions is of the form

$$P(s, L) = s^{-\tau_s} \mathcal{F}_s \left(\frac{s}{s_c(L)} \right), \tag{6.19}$$

$$P(t, L) = t^{-\tau_t} \mathcal{F}_t \left(\frac{t}{t_c(L)} \right), \tag{6.20}$$

where $\mathcal{F}_x(z)$ are scaling functions that are approximately constant for $z < 1$, and decay very fast to zero for $z > 1$. The quantities $s_c(L)$ and $t_c(L)$ are the cut-off characteristics size and time, which are assume to depend on system size as $s_c(L) \sim L^D$ and $t_c(L) \sim L^z$, thus defining the standard critical exponents τ_s, τ_t, D (the fractal dimension) and z (the dynamic critical exponent) (Pruessner, 2012).

Assuming the scaling form given by Equations 6.19 and 6.20, we can compute numerically the associated critical exponents applying the moment analysis technique (De Menech et al., 1998). One starts by defining the q-th moment of the avalanche size distribution on a box of size L as

$$\begin{aligned} \langle s^q \rangle_L &= \sum_s s^q \, P(s, L) \simeq \int ds \, s^{-\tau_s + q} \mathcal{F}_s \left(\frac{s}{L^D} \right) \\ &= L^{D(q+1-\tau_s)} \int dx \, x^{-\tau_s + q} \mathcal{F}_s(x) \sim L^{\sigma_s(q)}, \end{aligned} \tag{6.21}$$

where we have introduce the FSS form in Eq. (6.19), and taken the continuous approximation for the s variables. The exponents $\sigma_s(q) \equiv D(q + 1 - \tau_s)$ can be estimated as the slope of the numerical evaluation of $s^q{}_L$ as a function of L in a double logarithmic plot. Then, for sufficiently large values of q, we can perform a linear fit of the exponent $\sigma_s(q)$ to the form

$$\sigma_s(q) = Aq + B, \tag{6.22}$$

with $A = D$ and $B = D(1 - \tau_s)$, from where D and τ_s can be directly estimated. Along the same lines, the exponents associated to the avalanche time distribution can be evaluated considering the q-th moment of the time distribution, $\langle t^q \rangle_L \sim L^{\sigma_t(q)}$,

with $\sigma_t(q) \equiv z(q+1-\tau_t)$.

Application of this method leads to the characteristics exponents reported in Table 6.2. We check the accuracy of these values performing a data collapse analogous to that performed for the avalanches in real fish, which, for the cumulated distributions, takes the form,

$$L^{D(\tau_s-1)}P_c(s) = F'_s(sL^D), \qquad L^{z(\tau_t-1)}P_c(t) = F'_t(tL^z), \tag{6.23}$$

The collapsed cumulated distributions are shown in Figure6.18(A) for avalanche sizes and Figure6.18(B) for avalanche durations. As we can see from these values, the exponents show a dependence on the value of the noise η, although the size exponents appear to be compatible within error bars.

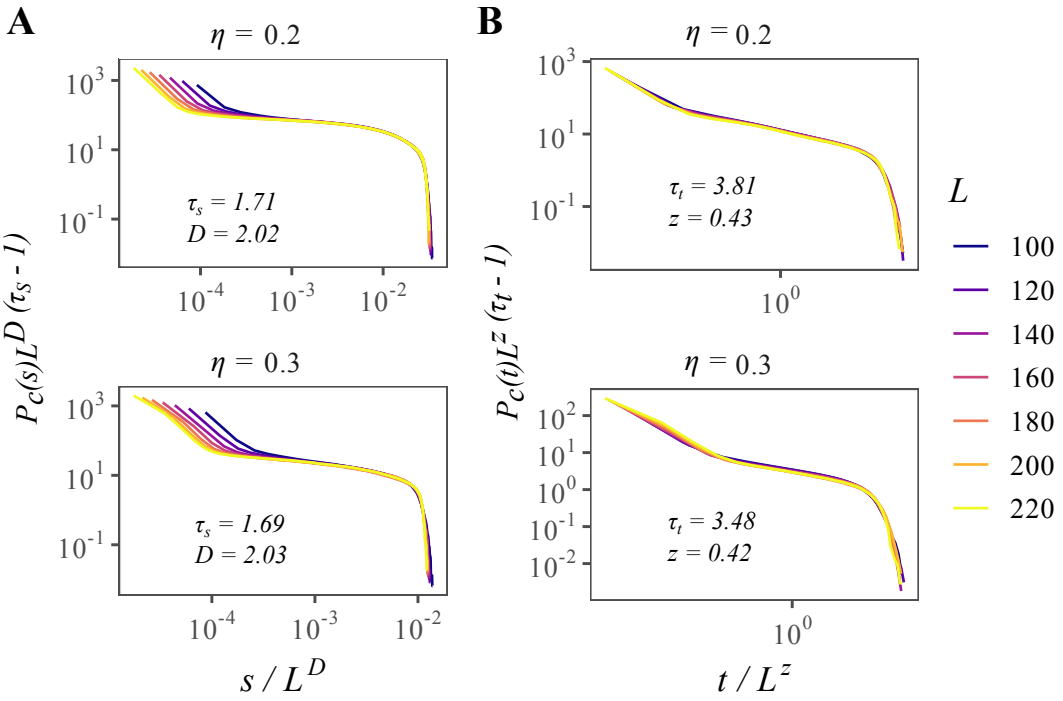


Figure 6.18: **A**: Check of the scaling of the avalanches cumulated size and duration distributions as given by Eq.(6.12) in a system with $\eta = 0.2$. **B**: Check of the scaling of the avalanches cumulated size and duration distributions in a system with $\eta = 0.3$. Statistics are performed over at least 10^5 different avalanches.

We have finally checked the effects of changing the turning threshold in the scaling of the distributions as a function of the system size. In Figure6.19 we show the

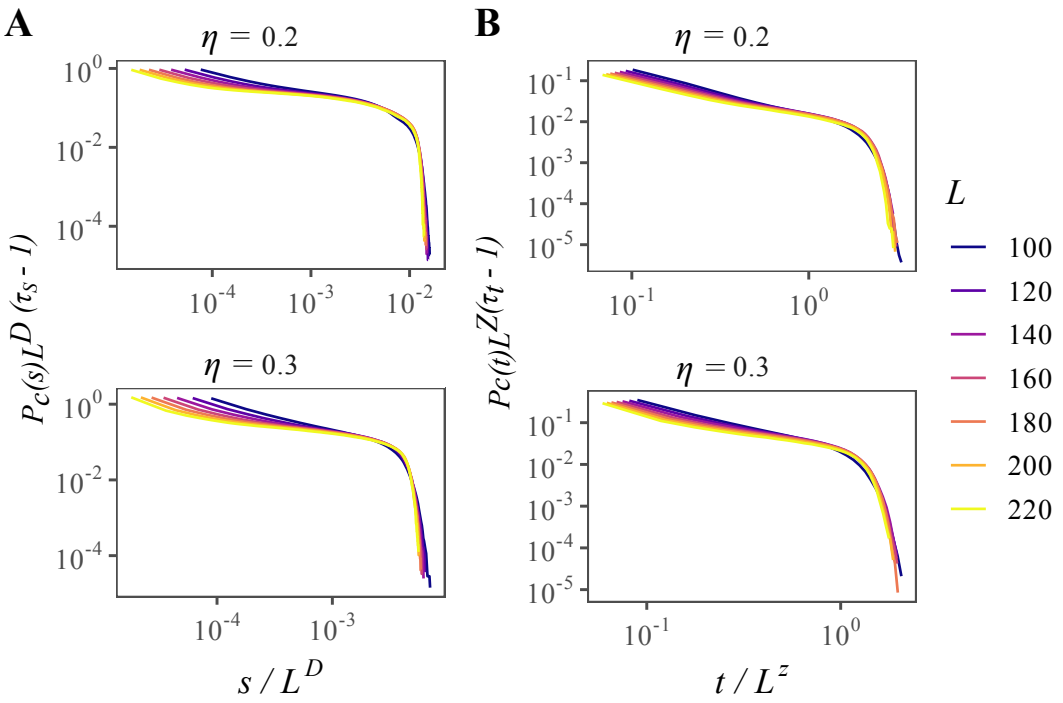


Figure 6.19: **A**: Check of the scaling of the avalanches cumulated size distributions as given by Eq.(6.12) in a system with $\eta = 0.2$ and $\eta = 0.3$, and turning threshold $\varphi_{th}(\eta) = 2.8\pi\eta$. **B**: Check of the scaling of the avalanches cumulated duration distributions in a system with $\eta = 0.2$ and $\eta = 0.3$, and turning threshold $\varphi_{th}(\eta) = 2.8\pi\eta$. Statistics are performed over at least 10^5 different avalanches.

results for a turning threshold $\varphi_{th} = 2.8\pi\eta$ for $\eta = 0.2$ and 0.3, summarized in Table 6.2. As we can see, the scaling exponent τ_s and τ_t in our model depend on the value of the threshold. This fact is in contrast with the behavior of the fish school, in which the characteristics exponents appear to be independent of the threshold, and thus allowing for a scaling solution of the form given by Eq.(6.11). Interestingly, the exponents D and z appear to be rather detail independent, taking the approximate values $D \simeq 2$ and $z \simeq 1/2$ for any value of η and φ_{th}, which would indicate that avalanches in this model are compact (Pruessner, 2012).

It is important to notice that the presence of a rotating leader in necessary to obtain scaling avalanche distributions. Even in the absence of a leader, the heading fluctuations due to noise and interactions in the standard Vicsek model allow to define avalanches for a given threshold. These avalanches, however, show a simple, short ranged exponential distribution, as shown in Figure6.20.

Vicsek model with a perturbed leader $\varphi_{\text{th}}(\eta) = 2.5\pi\eta$				
η	τ_s	D	τ_t	z
0.2	1.73(5)	2.01(2)	4.03(5)	0.39(5)
0.3	1.69(5)	2.01(2)	3.49(5)	0.42(5)
Vicsek model with a perturbed leader $\varphi_{\text{th}}(\eta) = 2.8\pi\eta$				
η	τ_s	D	τ_t	z
0.2	0.99(5)	2.06(2)	0.26(4)	0.50(5)
0.3	1.04(5)	2.03(2)	0.57(5)	0.52(5)

Table 6.2: Summary of scaling exponents for the avalanche size and duration distributions obtained from the Vicsek model with a perturbed leader.

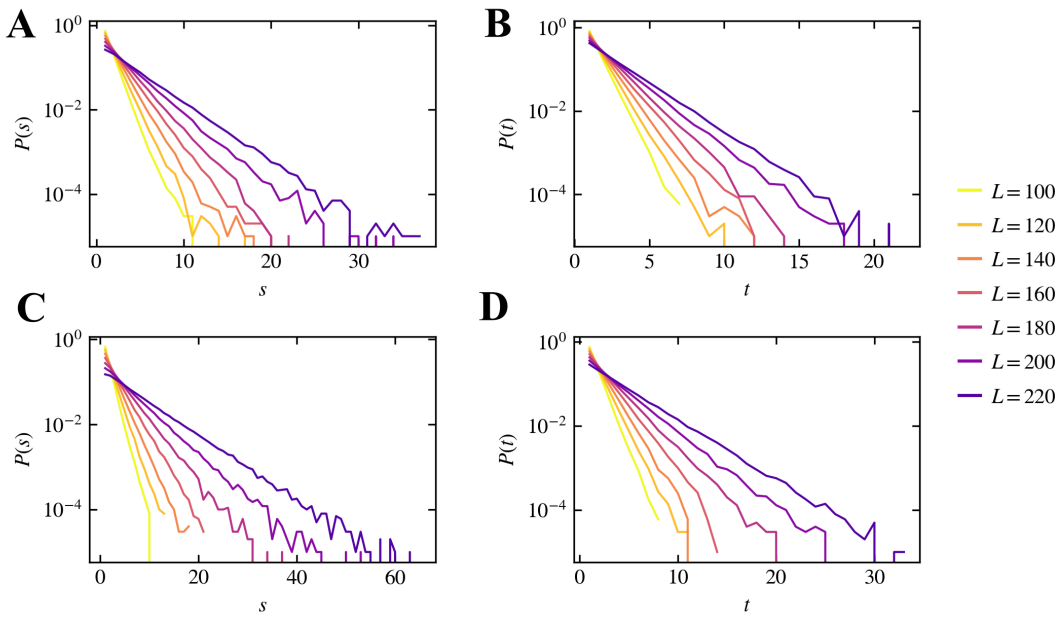


Figure 6.20: **A**: Cumulative probability distribution of size $P_c(s)$ of avalanches in the standard Vicsek model with $\eta = 0.2$, turning threshold $\varphi_{th}(\eta)$ and different values of L. **B**: Cumulative probability distribution of duration $P_c(t)$ in the standard Vicsek model with $\eta = 0.2$, turning threshold $\varphi_{th}(\eta)$ and different values of L. **C**: Cumulative probability distribution of size $P_c(s)$ of avalanches in the standard Vicsek model with $\eta = 0.3$, turning threshold $\varphi_{th}(\eta)$ and different values of L. **D**: Cumulative probability distribution of duration $P_c(t)$ in the standard Vicsek model with $\eta = 0.3$, turning threshold $\varphi_{th}(\eta)$ and different values of L. Statistics are performed over at least 105 different avalanches.

6.3 Discussion

Behavioral cascades, taking the form of intermittent rearrangements (avalanches) in the patterns of movement are an important, albeit sometimes neglected, feature of collective motion in animals. Here we have shown that behavioral cascades can be observed in the rearrangement dynamics of swimming fish schools. Such avalanches, defined in terms of a turning threshold for the heading of the fish, have

distributions of sizes and times exhibiting a scaling behavior compatible with a power-law tail truncated by a cut-off that is an decreasing function of the turning threshold. A data collapse analysis allows to determine the exponents characterizing the scaling form. We conjecture that such avalanche behavior can be due to the presence of effective leadership in the schools. In order to support this conjecture, we introduce a measure of leadership, based in the concept of avalanche initiators, and observe that, indeed, some fish have consistently an unusually large probability to initiate any avalanche in which they participate. These predominant initiators can be interpreted as effective leaders, determining the start of sudden rearrangements of the school headings. Leadership in the context of avalanche initiation could account for individuals having sudden behavioral changes or specific information about the environment, such as the proximity of a wall.

To check whether the presence of leaders is enough to induce avalanche behavior in collective motion, we have considered a very simple model, consisting in a variation of the classical Vicsek model with the addition of a global leader, which influences the movement of the particles, subject to random changes in its heading. Interestingly, this simple model displays an intermittent behavior qualitatively similar to that observed in real schools, with avalanche size and duration distributions displaying a self-similar scaling form.

Our results provide a new perspective on the avalanche behavior observed in real collective motion situations (Ginelli et al., 2015), which can be associated to a simple mechanism of leadership, observed in many natural situations, indicating the possibility of a direct relation between these two phenomena. Leadership in the present context of a moving school corresponds to those individuals that first react to any external input, or that first exhibit a random behavioral change, and preferentially start sudden rearrangements of the trajectories of other fish in the school. Such interpretation is validated by the numerical results from our model. It is also worth emphasizing that, while it does not offer a perfect quantitative prediction of the characteristic exponents, it nevertheless allows to reproduce the scaling form of the avalanche distributions within a minimalist modeling framework.

Different venues of future research stem from the results presented here. From an empirical perspective, it would be interesting to further study the nature of the

avalanches observed in real schools, and to correlate them with other physical properties measured in similar systems (Herbert-Read et al., 2011; Katz et al., 2011), as well as with other measures of leadership devised in different contexts of collective motion (Strandburg-Peshkin et al., 2018). From a numerical point of view, our results present new challenges in the understanding of the properties of the proposed model. Indeed, a clearly open question remains to ascertain the ultimate origin of the scaling behavior observed in avalanches in a system in which no apparent critical transition exists. Another interesting question regards the effects of leader switching strategies. We expect the scale-free nature of the observed avalanches to be preserved, provided that the influence of the leader, sensory wise, remains rather long-ranged. In this sense, as we have numerically checked (data not shown), a short-ranged leader, only with local influence over its nearest neighbors, is not able to induce system-wide orientation rearrangements. On the other hand, the value of the exponents associated to the size and duration cutoffs are apparently independent of the noise intensity imposed on the system. These observations hint towards a possible partial universality, which is not shared, however, by the power-law decay exponent. Further work in this direction is clearly needed in order to clarify these issues.

6.4 Closing remarks

In our empirical analysis we found that the measured avalanches distributions of size and duration of the measured avalanches show a scale-free behavior, in analogy with self-organized critical processes (Pruessner, 2012), that can be described in terms of a set of characteristics scaling exponents. A statistical analysis of avalanche initiators allowed us to differentiate fish with an anomalous large probability of starting an avalanche, acting thus as effective leaders promoting substantial school rearrangements. By including a global leader that periodically changes direction in the simplest SPP model, the Vicsek model, we were able to replicate the intermittent scale-free avalanche-like behavior, not present in the original model.

Our results confirm the presence of scale-free signatures in behavioral cascades

in collective motion (Rosenthal et al., 2015) and highlight the role of effective leadership interactions in the emergence of this sort of collective behavior.

Part 2

Collective-decision making in swarming robots

7 Kilobots: swarm robots to study collective decision making

7.1 What are kilobots?

Kilobots are simple open-source swarm robots measuring 3.3 cm in diameter especially designed for testing collective behavior. They were developed by the Self-organizing Systems Research Group at Harvard University (Rubenstein et al., 2012) and are now manufactured by the Switzerland based company *K-team*. The kilobot was designed with the capability to perform a wide variety of collective behaviours, but is not so complex that it is unaffordable (the cost per kilobot is $\approx$ 100 euros).

They have been used to study collective decision-making behaviors (Valentini et al., 2016), pattern formation (Gauci et al., 2018), morphogenesis (Slavkov et al., 2018), space exploration (Dimidov et al., 2016), and collective transport of objects (Rubenstein et al., 2013).

Each individual bot consists of an Atmega328 microprocessor with 32 KB flash memory and 1 KB EEPROM. Locomotion occurs from two lateral coin-shaped vibration motors which cause the supporting legs to vibrate, therefore allowing the bots to move forward when both motors are activated, and rotate to each side by activating only the left or right motor. The bots move approximately 1 cm/s and rotate approximately 45°/s. This type of locomotion does not allow for odometry (i.e. position estimation relative to some known position),which limits precision when moving over long distances. The kilobot has a 3.4 V 260 MAh lithium-ion battery, which can potentially power the robot for up to 24 hours, depending on its activity level. Figure 7.1 shows the main components of kilobots from below and isometric views.

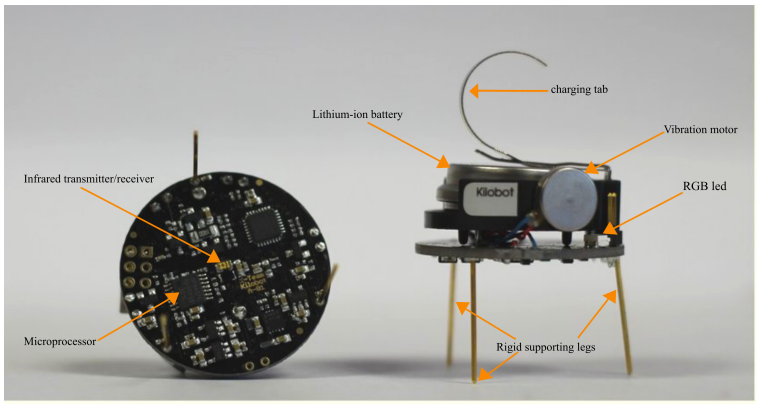


Figure 7.1: Kilobot architecture and main components from below and isometric views.

Kilobots have the capacity to exchange up to 9 bytes of information with nearby robots within an interaction radius. This information is communicated by an infrared LED transmitter and an infrared photo-diode receiver, located at the center of the circuit board (Figures 7.1 and 7.2), with an isotropic emission or reception pattern, which allows each robot to receive messages equally from all directions. The kilobots can communicate at rates of 30 kb/s with robots up to 10 cm away, although this distance may vary slightly depending on a few factors, such as reflective characteristics of the surface and the luminance of the ambient light (Rubenstein et al., 2012). To deal with potential communication "jamming" of two or more robots transmitting at the same time, a standard *carrier sense multiple access with collision avoidance* (CSMA/CA) protocol was implemented (Karn et al., 1990); this allows for multiple signals to use the same channel by waiting until it is free. However, even with the implementation of CSMA/CA the channel can get saturated when many kilobots are transmitting simultaneously.

When communicating, the receiving robot measures the intensity of the incoming infrared light using two amplifiers, allowing it to calculate the distance between neighbours. Distance measurements, then, are used to gauge the relative position and orientation of nearby robots. This helps each bot self-locate. Due to noise and manufacturing variances, the intensity of the incoming light has an accuracy of $\pm 2mm$. In addition, an RGB LED light display on the kilobots shows information about their state to external observers.

Kilobots are programmed in C-language with specific functions handled by a

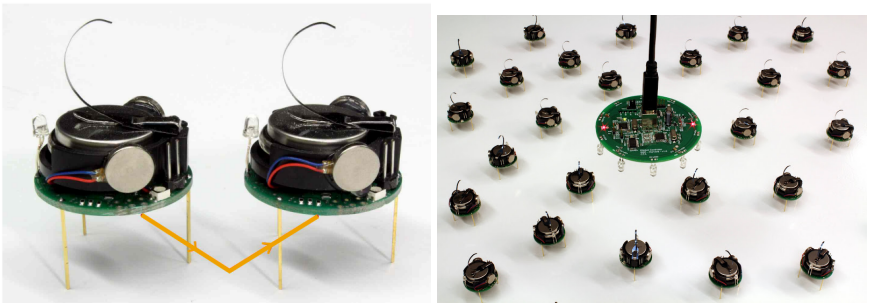


Figure 7.2: textbfA: Example of a kilobot communicating to another via reflected infrared. **B**: Overhead controller (OHC) sending infrared signals to multiple kilobots at the same time.

custom Application Programming Interface (API) packed as a library named *kilolib* (*Kilolib* n.d.). Kilobots have their own clock, the *kilotick*, which is equivalent to approximately 30 milliseconds. Additionally, time can be measured by the number of iterations per time unit. Programs are loaded on the kilobots by an overhead controller (OHC) that sends the program via infrared to all kilobots synchronously, making the system scalable and avoiding the time-consuming task of programming each one separately. The OHC also connects with low-level electronics like motors, power circuitry and RGB LED to communicate information about them to the user.

To upload and run the programs a Graphical User Interface is used, the *KiloGUI*, transmits directly to the OHC, which in turn communicates with the kilobots. Figure 7.3 show the main GUI window, where the user can access several attributes of the kilobots such as to upload, reset, pause or run a program, check kilobots batteries' voltage, activate sleep mode, and access messages sent by the kilobot (*Serial Input*) and calibrate left and right motors power (*Calibration*).

The *Serial Input* button opens a new window where all the messages sent by a robot appear. This option serves as a debugging mechanism because the user can get direct feedback from the kilobot. To do so, the kilobot has to be directly connected to the OHC by a debug 2-wire cable, respecting the polarity as seen in Figure 7.4.

7.2 Experimental characterization of individual kilobots

Due to differences in manufacturing, motion accuracy varies based on the individual kilobots and on the surface upon which they perform their motion and communication. Before implementing our behavioral algorithm for the study of collective

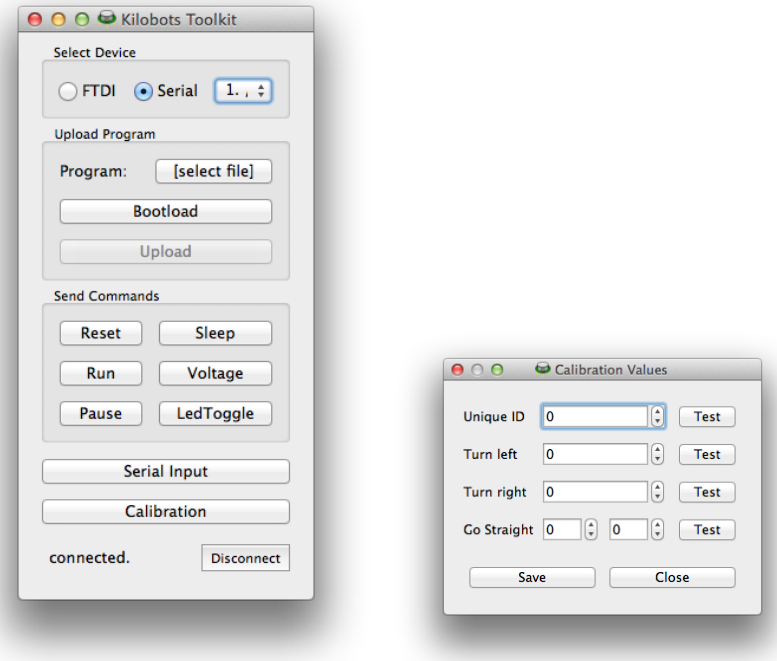


Figure 7.3: **A**: KiloGUI, the kilobots graphical user interface (GUI) to control the kilobots through the OHC. **B**: Calibration GUI to manually calibrate motors power values for turning right, turning left and going straight. Here also a unique identifier (UID) can be assigned to each kilobot.

decision making, we characterized individual speed and message reception of individual robots in our experimental setup, consisting either on a plain $1m^2$ bounded melamine whiteboard where kilobots are placed, or on a glass surface held 15 cm from the white melamine base. The second setup was used to guarantee the synchronization of the kilobots' internal clocks as will be explained in detail in Chapter

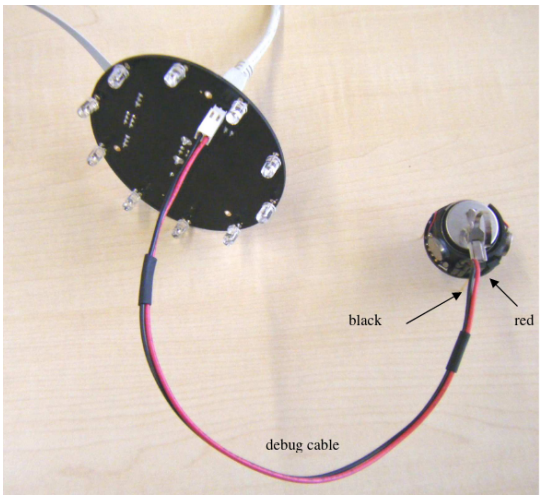


Figure 7.4: Debugging configuration. Image borrowed from Tharin, n.d.

9.

As described in Section 7.1, kilobots forward motion and turning are produced by lateral vibrating motors. The power of the motors is calibrated to achieve a straight line when going forward and a full left and right turn. Despite being calibrated by the manufacturer, we found that most deviated significantly from a straight line moving forward. With this consideration, we manually calibrated each of the kilobots' motors power values to achieve full left and right turns and, as much as possible given their stick-slip locomotion, a straight forward motion.

To calibrate the motors, we used the KiloGUI *Calibration* option in Figure 7.3 A, which opens a new GUI window to adjust the power values for turning left, right and going straight, as shown in Figure 7.3 B. Adjusted values are transmitted through the OHC and the motors' power is modified in real time. The values of each kilobot varied from 55 to 95, with left and right values to go straight were 2 to 10 units smaller than the turning values. Once the motors were calibrated, we wrote the values changes to the EEPROM memory of the individual kilobot.

In the same calibration window, one can modify the bot's ID. We assigned a unique identifier (UID) to each bot, allowing them to be recognized by others as the same individual while performing the collective decision making process.

According to Rubenstein et al., 2012 kilobots move straight at an average speed of 1 cm/s. To build upon this assessment, and to evaluate the speed distribution of our kilobot sample, we measured individual speed v_i by timing the seconds that a bot moved 15 cm in *straight line*, and then converted it to cm/s. We documented mean and standard deviation for reproducibility from our physical kilobots to a kilobot simulator, described next in 7.3. To increase insight of accuracy within our sample size, we repeated the measurement 10 times per kilobot, as there is always some small variation in their displacement. Figure 7.5 shows the v_i distribution. Kilobots move at a mean value $\bar{v}_i = 0.90 \pm 0.2 cm/s$, with some bots moving faster, with $v_{max} = 1.49$ cm/s and some moving as slow as $v_{min} = 0.43$ cm/s. Within our experiment, our kilobots' mean speed aligns with the mean speed documented by Rubenstein et al., 2012. However, there is a considerable amount of dispersion in speed values that must be taken into account when understanding their dynamics and testing our algorithms.

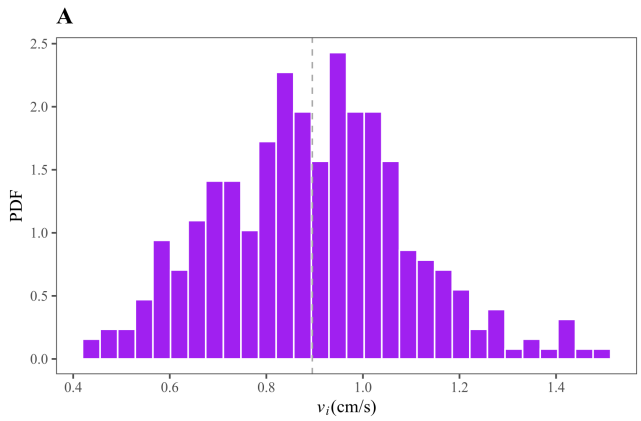


Figure 7.5: PDF of individual speed v_i (cm/s) for 35 kilobots. Each kilobot speed was measured 10 times, adding up to a total of 350 values.

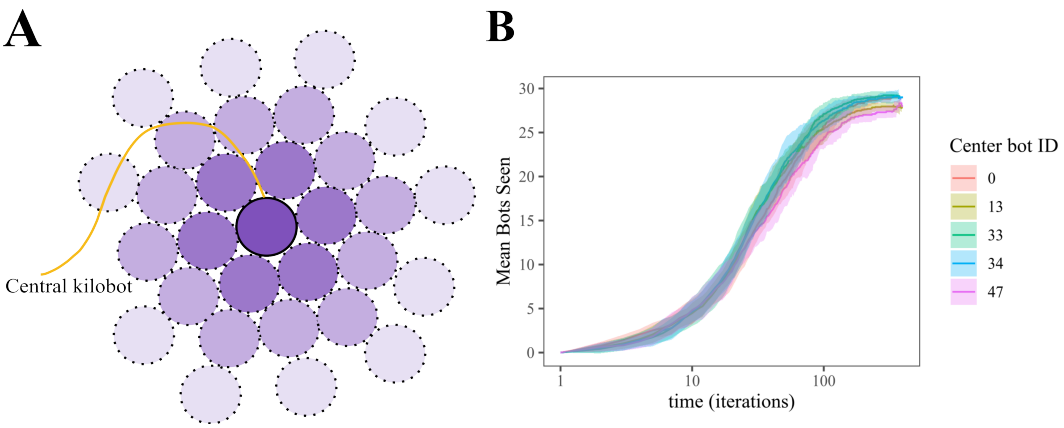


Figure 7.6: **A**: Bots seen by central bot debugging configuration. **B**: Mean bots seen by central kilobots as a function of time (iterations). Colors represent different central bots and shaded regions the standard deviation from the mean.

To characterize how much time it takes for a focal bot to receive messages from neighboring kilobots, we worked with 30 kilobots. We placed one kilobot at the center and surrounded it with the rest organized in three kilobot layers, i.e. with the farthest kilobots located at three bot-size distance ($\sim$ 9 cm) from the central one, as shown in Figure 7.6 A. We then connected the debug cable to the central kilobot and the OHC, and obtained the number of time iterations that are needed for the central bot to detect the maximum number of bots that are within its communication range. We repeated this measurement thirty times per kilobot, for five different central kilobots. Bots see most surrounding bots at around 100 loop iterations ($\sim 1s$), as shown in Figure 7.6 B.

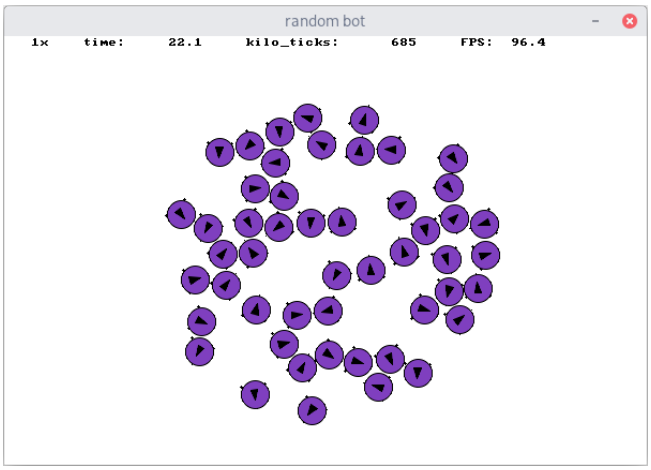


Figure 7.7: Kilombo GUI shows kilombo bots performing the same algorithm as physical kilobots.

7.3 Kilombo: a kilobot simulator

Collective algorithms testing and debugging in the actual physical kilobots is a challenging task. Setting up and programming the kilobots is both time intensive and meticulous in its process, and the time needed for the behavioral algorithm to run is often extensive. Moreover, unexpected behaviors or errors are difficult to deduce only by the kilobots behavior. Due to these elements, it is useful to test algorithms in a computer simulator prior to executing physical experiments. Simulations perform algorithms faster and allow for inspection and repetition when a problem is detected.

With this in mind, we tested our behavioral algorithms on a simulator designed specially for kilobots. Kilombo is a C-based simulator that allows the simulation code to be run on the physical robots as well, removing the slow and error prone step of converting code to a different platform (Jansson et al., 2015). The simulator also includes a graphical user interface (GUI) that shows the robots from a top-down perspective as in Figure 7.7 and allows them to interact with the simulation at run time by moving robots around with the mouse, or changing the simulation speed. The GUI can be deactivated.

Kilombo simulations capture the essential features of the physical kilobots and the real world they interact with that can affect their behavior. The simulator keeps track of the Euclidean positions and orientations of each robot, and updates these through the simulation at each time step (at each time step the user program's loop

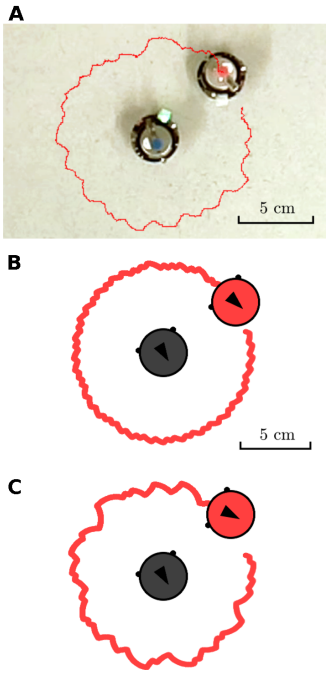


Figure 7.8: Kilombo demonstration by Jansson et al., 2015 of parameter tuning to represent the physical kilobots behavior more accurately. **A**: Physical kilobots performing the *orbit* algorithm.**B**: Bots in kilombo performing the same *orbit* algorithm. **C**:Bots in kilombo performing the same *orbit* algorithm but adding noise to message reception so that 20 % of the messages are lost, and a standard deviation of 2 mm

function is run). Motion occurs as in the real kilobots, by turning the two motors on or off. A configuration file allows for setting bots speed and turning rate, so they can reflect the physical robots' real motion depending on the surface used and the robots' motor calibration. The simulator also handles collisions between robots by displacing overlapping robots equally along the vector joining their centers with no energy loss to friction or plastic deformation. Furthermore, the simulator replicates communication and the distance measurement via infrared, including variance in messages reception success and distance measurement. For example, Figure 7.8 shows a comparison of kilombo performing the kilobots' sample algorithm *orbit* (Rubenstein et al., 2012) which consists of a stationary bot that emits messages. Another bot moves and tries to keep a constant distance to the stationary bot, by alternating between turning left and right. Figure 7.8 A shows the moving bot trajectory in physical kilobots and Figure 7.8 B shows the same algorithm implemented in kilombo. Figure 7.8 C shows the same algorithm but modifying the noise in message reception and distance measurement.

The user can also include a callback function to specify a light intensity profile

Parameter	Description (units)	Value
Number of bots	Number of bots to simulate	35
Time step	Simulator time step (seconds per loop).	0.0103
Simulation time	Total simulation runtime (s).	872
Turn rate	Kilobot turning speed (degrees/s)	45
Speed	kilobot movement speed (mm/s)	9
Speed variation	Standard deviation in kilobots' movement speed (mm/s)	2

Table 7.1: Parameters chosen in kilombo to accurately represent physical kilobots. Other parameters not mentioned in the table are left equal to kilombo presetting values.

for the environment for kilobots' light sensing, and another callback function for user-defined obstacles that limit kilobots motion through space. Other features that can be configured in the simulator include initial bots space distribution, random seed generation, and seconds per time step for real world compatibility. For further details and information the Kilombo User Manual can be accessed online [1].

Kilombo has proved to be sufficiently accurate for testing potential behavioral algorithms (Jansson et al., 2015), although, as with any simulator, it is difficult to predict to which extent small variations of the real world can propagate through the system and affect global dynamics. Experimentation with real kilobots is then fundamental to complement kilombo simulations. Furthermore, kilobots behavior can be contrasted to what is observed in the simulator to help identify algorithms that are more sensitive to the noise of real life (Jakobi, 1997).

7.4 Kilobots motion characterization in kilombo simulations

We will now take profit from the kilombo simulator to characterize the collective motion of kilobots. The simulator has the advantage to enable the tracking of individuals' positions over time. Therefore, after having characterized the kilobots' individual speed distribution in Section 7.2, we feed the obtained mean velocity and standard deviation to the kilombo configuration file. Table 7.1 shows the full set of configuration parameters.

We are interested in characterizing the kilobots collective motion while bots perform a continuous time persistent random walk (PRW), i.e. as they alternate a forward motion at a constant speed with random changes of its direction of motion

[1] https://github.com/JIC-CSB/kilombo/blob/master/doc/manual.md

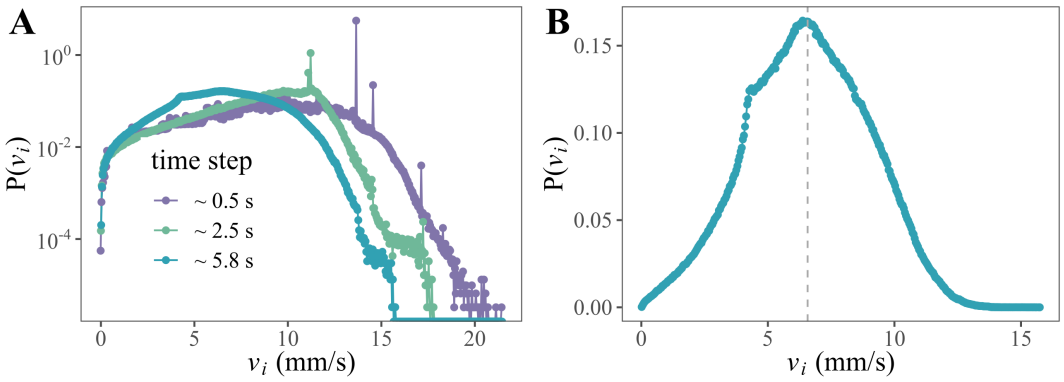


Figure 7.9: A: Probability distribution of the mean individual speed $\bar{v}_i$ (mm/s), calculated from bots' positional data every $\Delta t \simeq 0.5$, 2.5 and 5.8 s, shown in log-lin. **B**: Probability distribution of the mean individual speed $\bar{v}_i$ (mm/s) for $\Delta t = 5.8$ s in a linear plot.

with a prescribed turning rate, in a circular arena of radius $R = 20$ cm. There are many examples of non-equilibrium self-propelling entities in nature for which this model would be appropriate. Examples include self-propelling living entities at different length scales, from bacteria to fish schools. (Ariel and Ayali, 2015; Murakami et al., 2015).

We conducted a study involving $N = 35$ kilobots (to match the number of physical kilobots available) for a total of 25 runs, each lasting 872 seconds (equivalent to 14.5 minutes). Individual bot's positions were saved every $\Delta t \sim 6$ loop iterations. To prevent, as much as possible, kilobots from clustering at the border of the arena, the PRW was configured with discrete wide turning angles so they would be able to turn away from the border faster. We established the turning times to either ~ 2.8 (turning angle of ~ 125 degrees) or ~ 5.8 seconds (turning angle of ~ 251 degrees), and a forward motion time of ~ 3.8 seconds. From these data, next we analyze the collective motion and distribution of the kilobots within the circular arena.

7.4.1 Speed distribution during kilobots PRW motion

By taking the modulus of the individual bots displacements in a given time window (Δt), we can compute their corresponding speed. We have taken each bot positional data and estimated their individual speed v_i (mm/s) as the displacement in the time window divided Δt.

In Figure 7.9 A, we present the probability distribution of such mean individual speeds calculated over different time windows. For short Δt, the histogram reveals

notably larger speed values and characteristic "peaks" at particular v_i values. Those features tend to disappear as Δt is increased. The rational for this behavior of $P(v_i)$ is the following: First, let us recall that kilobots are performing a PRW, with phases of straight movement and phases of rotation to change their direction of motion. In this regard, we have set for the kilombo bots' mean speed and standard deviation parameters the speed measured in the physical kilobots when moving in a straight fashion and independently, $\bar{v} = 9 \pm 2$ mm/s, and we have fixed the turning speed to $45°/s$, based on prior work by Rubenstein et al., 2013 (Section 7.2).

Second, it turns out that, while a kilobot rotates, its mean (translation) speed calculated over a short time window (shorter than the rotation duration) is higher than when they move straight. Of course, this does not mean that the bot is actually moving faster. What happens is that its rotation is not a pure rotation around one of its legs: a kilobot rotates by activating only one of its lateral vibrating motors, which causes its body to rotate around an axis that slightly moves in time (see, for instance, Figure 7.8). The vibrational drift of the rotation axis appears to have an associated characteristic speed, which can be clearly identified as very narrow peaks in the speed distribution. Thus, these narrow peaks appear as a consequence of the erratic characteristics of the kilobot vibrating motion on relatively short length-scales, but they become irrelevant on the longer timescales involved in their PRW dynamics. Using a longer time window, $\Delta t = 5.6$ s (already slightly larger than the typical rotation stage duration), such rotation-induced peaks are less relevant and the distribution curves smooth out.

Still, as seen in Figure 7.9 B, the individual velocities of the group of kilobots moving concurrently in the arena are found to deviate from the pure translation input value parameter $\bar{v} = 9 \pm 2mm/s$ used in simulations. These deviations can indeed be attributed to the periods of time spent by the bots reorienting their direction of motion during the turning stage of their PRW motion (which paradoxically can produce larger instantaneous speeds, but globally slow down the displacement), but also to collisions among bots and with the arena wall. The distribution in Figure 7.9 B shows that the effective mean individual bot velocity calculated over 5.6 seconds is equal to $\bar{v}_i = 6.50 \pm 2.48$ mm/s.

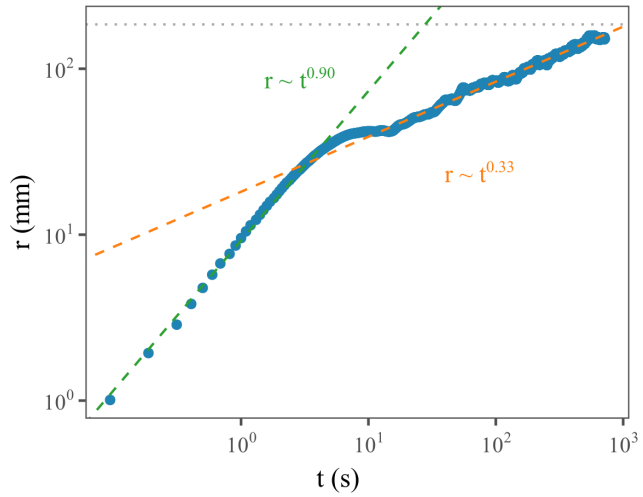


Figure 7.10: Root-mean square displacement $r(t)$ for $N = 35$ bots in kilombo. As a guide to the eye, dashed lines in green and orange mark the extension of the ballistic and subdiffusive regimes, respectively, observed in the dynamics. The grey dotted line shows the arena radius, i.e. the maximum distance a bot can travel.

7.4.2 Mean Square displacement

We now turn our attention to the characterization of the mean-square displacement of the simulated kilobots on the arena. We calculated the mean square displacement $r^2(t)$ as a function of the elapsed time t, averaged over N bots and over all time lags of duration t in the interval $[0, t_{max}]$,

$$\langle r^2 \rangle = \frac{1}{T-t} \frac{1}{N} \sum_{t_0=0}^{T-t-1} \sum_{i=1}^{N} [r_i(t_0 + t) - r_i(t_0)]^2 \tag{7.1}$$

where $r_i(t)$ is the position of kilobot i at time t.

Our results show that the root-mean square displacement is described by a crossover between two power-law-like behaviors of the form: $r = \langle r^2(t) \rangle^{1/2} \propto t^{\alpha}$, followed by a saturation plateau due to the finite size of our arena. We found that the exponent values for our PRW kilobots are equal to $\alpha \simeq 1$ at short times, and $\alpha \simeq 0.33$ at intermediate time scales, as illustrated in Figure 7.10. Thus, we conclude that, as expected, after an almost ballistic regime at short time scales, describing their forward motion, kilobots exhibit a weak subdiffusive behavior at longer time scales, when they perform their discrete PRW dynamics, due to collisions with other kilobots. Indeed, subdiffusive behavior is a characteristic feature of crowded systems that can limit the diffusion capabilities of individual bots.

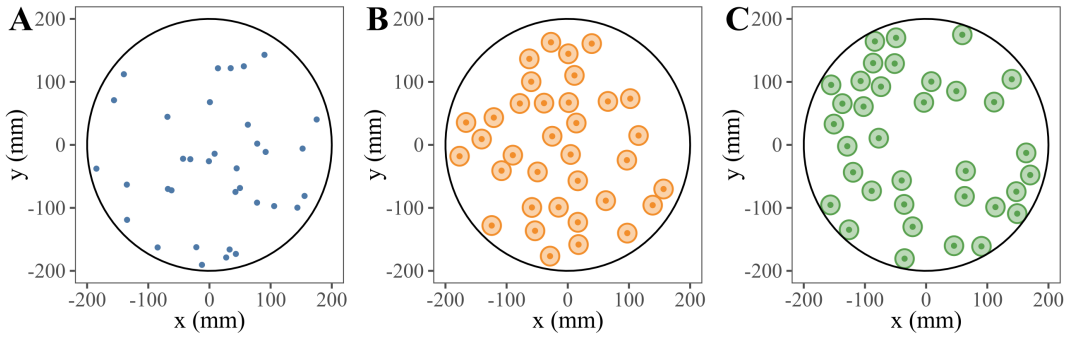


Figure 7.11: Examples of 35 bots displayed in an arena of radius $R = 200$ mm with **A**: no exclusion radius, $r_e = 0.0$ (point-like particles), **B**: an exclusion radius $r_e = 1.5$ cm, and **C** kilombo bots with an exclusion radius $r_e = 1.5$ cm.

7.4.3 Spatial distribution of kilobots inside the arena

As the kilobots move within the circular arena in our kilombo simulations, we further analyze their spatial distribution. In order to better understand the unique characteristics of the kilobots' spatial coverage due to their particular dynamics, we also examine the results obtained from fixed (quenched) configurations of randomly positioned bots in the same arena. This is done with the purpose of better evaluating the deviations from the quenched approximation (used as a reference model) observed in the case of moving bots.

Thus, each spatial configuration in the quenched approximation is an independent sample of the positions of N agents inside a circular arena of radius R. In particular, we randomly place $N = 35$ agents within an arena of radius $R = 20$ cm. An agent position is generated from a uniform distribution inside a circle,

$$r = R\sqrt{\xi_1(0,1)}, \quad \theta = 2\pi\xi_2(0,1) \;\rightarrow\; x = r\cos\theta, \quad y = r\sin\theta \tag{7.2}$$

where $\xi_i(0,1)$ are random numbers uniformly distributed between 0 and 1. Agents have an exclusion radius of r_e due to its finite size, and occupy an area equal to πr_e^2. Therefore if the random deposition of a new agent overlaps with the area of an agent already placed in the arena, we discard this position and generate a new one. We analyze both the cases of finite-size bots with ($r_e > 0$) and zero-body size bots ($r_e = 0$). If $r_e = 0$ (i.e. point-like particles), bots never overlap.

Figures 7.11 show example quenched configurations of point-like bots (A), finite-size bots with an exclusion radius $r_e = 1.5$ cm (B), and of kilobots in the kilombo

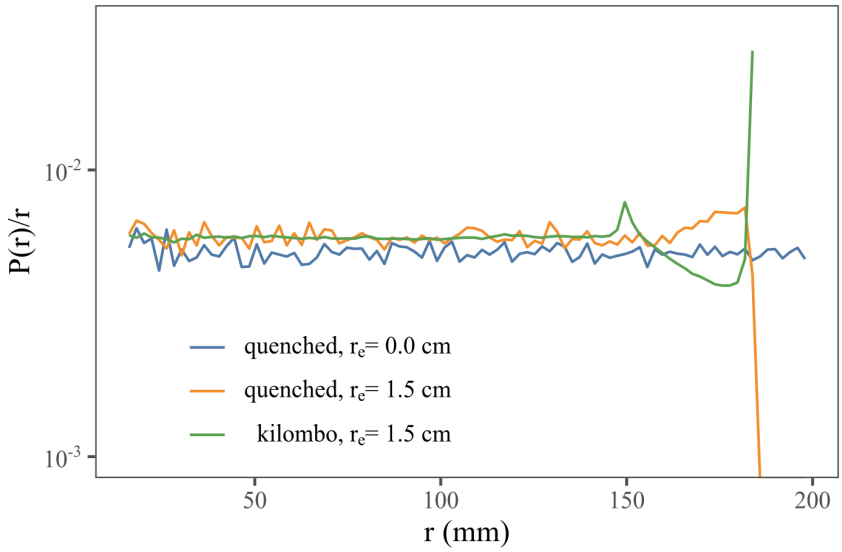


Figure 7.12: Normalized probability distribution of radial positions inside a circular region of radius r, $P(r)$, for quenched spatial configurations (with and without exclusion radius r_e), and for kilombo simulations.

simulator, which also have a radius $r_e = 1.5$ cm (C).

In the following subsections, we analyze the spatial distribution of kilobots within the kilombo environment, and compare it with the distribution of random quenched configurations. For this, we compute the probability distribution of radial positions of the bots, and the number of bots within circles of varying radii. Additionally, we examine fluctuations in the number of detected bots as the radial distance from the center of the arena is increased. By comparing our findings to quenched spatial configurations, and to theoretical predictions for point-like processes, we aim to better identify and describe any possible deviations from the expected outcomes for systems that lack excluded volume interactions or dynamical correlations effects.

Probability distribution of radial positions inside the area

In Figure 7.12, we illustrate the radial position probability distribution, or density profile, of kilobots in the steady state as a function of their radial distance to the origin of coordinates located at the arena center. This probability has been properly normalized by the area of a differential cylindrical shell between $(r, r + dr)$ to account for area differences at different radial distances. While point-like particles distribute uniformly throughout the entire arena, both the quenched approximation

with an exclusion radius and the simulated kilobots system exhibit noticeable deviations from this ideal behavior near the wall.

The characteristic dynamics of kilobots produces significant deviations from the random point-particle case, and gives rise to a prominent peak at $r \simeq 185$ mm where the probability of finding a bot is about one order of magnitude greater than that observed in the central area of the arena. The accumulation of some bots near the wall also causes a depletion of probability next to the wall, with the chance of finding other bots at less than a bot-diameter distance decreasing due to their own excluded volume.

Additionally, a smaller peak is visible at a radius compatible with a secondary ring of kilobots that are temporarily stalled against the bots trapped at the wall. The normalized distribution indicates a flat distribution for small and intermediate radii, with the border effect only becoming noticeable for $r > 140$ mm.

Number of bots and its spatial fluctuations

To gain further insight into our kilombo configurations, and to investigate the potential presence of clustering phenomena, we also determine the number of bots located within circular regions of increasing radius r.

We define $N(r)$ as the number of bots inside a circle of radius r measured from the arena center, which is nothing but the cumulative distribution of $P(r)$. For the case of quenched random configurations, we observe that the variation of $N(r)$ as a function of r obeys the following expression

$$N(r) = N\frac{r^2}{(R^2 - r_e^2)} \tag{7.3}$$

where R is the radius of the arena and r_e is the excluded radius of each bot.

This expression can be derived assuming that the probability of finding an agent inside an area of radius πr^2 follows a Poissonian distribution with probability p, given by

$$p = \frac{A(r)}{A_T} = \frac{r^2}{(R^2 - r_e^2)}. \tag{7.4}$$

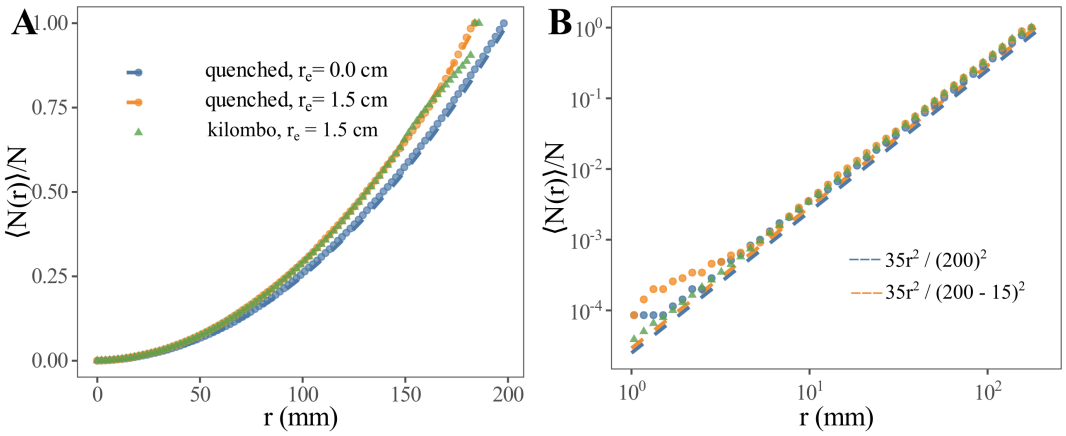


Figure 7.13: **A**: Fraction of bots $\langle N(r)\rangle/N$ as a function of r for kilombo simulations and for quenched bots' configurations. **B**: Log-log of $\langle N(r)\rangle/N$ as a function of r for kilombo simulations and for quenched bots' configurations.

Thus, $N(r)$ can be seen as a random variable that follows a binomial distribution with parameters N and p (Baddeley et al., 2007).

In Figure 7.13 A, we plot the time average $\langle N(r)\rangle/N$ as a function of r for quenched configurations and for kilombo simulation data. The dashed lines correspond to Equation 7.3 with $r_e = 0$ and $r_e = 1.5$ cm. $\langle N(r)\rangle$ grows essentially as the square of the radius of the containing circle r, as one would expect for a random homogeneous distribution of bots.

When agents have a finite body size $r_e \neq 0$, the effective area of the arena (where agents can be placed) is reduced to $\pi(R^2 - r_e^2)$. Moreover, finite body effects result in the arena filling up more quickly (as seen for both finite body cases), enhancing proximity between bots and, as we will discuss later on, creating bigger dynamic clusters of bots. Configurations obtained from kilombo simulations further show the border effect: $\langle N(r)\rangle/N$ slows down its growth right before the boundary and then suddenly rises. We also observe that the area πr_e^2 correction factor is necessary and works well, as in the quenched configurations, if we consider $R = 20$ cm and $r_e = 1.5$ cm.

The log-log plot in Figure 7.13 B shows $\langle N(r)\rangle/N$ growing as $\sim r^2$ for all the three cases. The discrepancies observed for small r are simply due to lack of statistics for the occurrence of observations at very small values of r.

Having analyzed the distribution of bots within a circular arena of radius r, we now proceed to also analyze its fluctuations. Remember that $\langle N(r)\rangle$ quantifies the

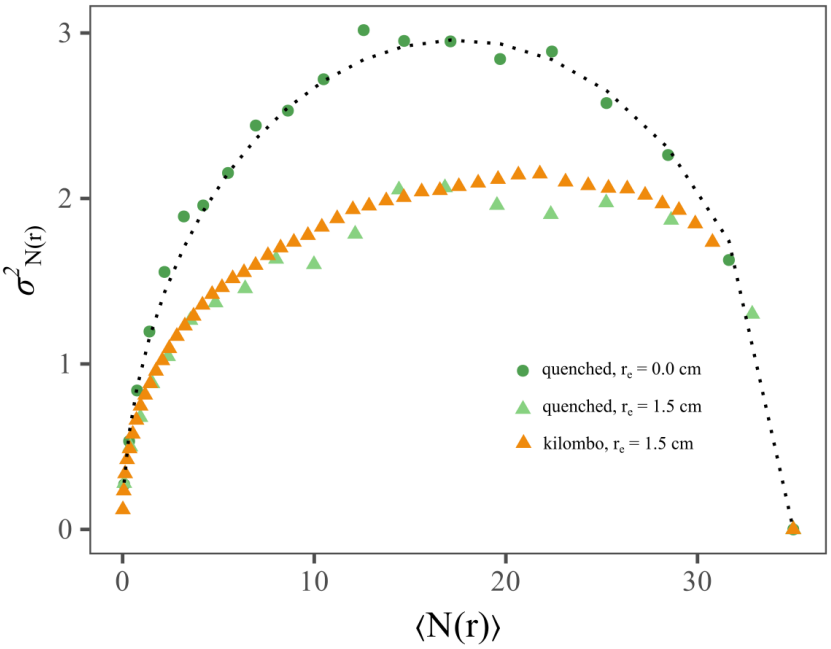


Figure 7.14: Mean-square number fluctuations $\sigma^2_{N(r)}$ as a function of the average number of bots $\langle N(r)\rangle$ in a circular area of radius r.

average number of particles in a circular region of radius r over different time steps t in the stationary state. Then, mean square fluctuations of $N(r)$ are calculated using the following expression,

$$\sigma^2_{N(r)} = \langle (N(r) - \langle N(r)\rangle)^2\rangle. \tag{7.5}$$

In a homogeneous case, with a uniform distribution of point-like bots in the arena of area $A = \pi R^2$, $\langle N(r)\rangle = Nr^2/R^2$, and the corresponding variance is

$$\langle N^2(r)\rangle - \langle N(r)\rangle^2 = \frac{Nr^2}{R^2}\left(1 - \frac{r^2}{R^2}\right). \tag{7.6}$$

This is the result for an ideal Poisson process of point-particles randomly located in the arena, where $N(r)$ has a binomial distribution with parameters N and $p = r^2/R^2$, as previously stated. In this particular case, fluctuations exhibit a symmetric inverted parabolic dependence on the average value $\langle N(r)\rangle$, and attain a maximum value at $N/2$.

From Figure 7.14 we can observe that, as expected, fluctuations in the case of point particles obey nearly perfectly the theoretical expression in Equation 7.6. On

the contrary, for finite-body size bots, $r_e \neq 0$, the behavior of fluctuations deviates from the theoretical ideal behavior, in a similar manner for both quenched configurations and kilombo simulations. The ideal theoretical parabolic behavior is now substituted by an asymmetric parabolic profile with enhanced fluctuations on its right hand side, i.e. for larger radial distances that contain more than half the total number of particles in the arena. In particular, its maximum value is now shifted towards larger values around $\langle N(r) \rangle / N \sim 0.6$ bots or, according to Figure 7.13, to radial distances beyond $r \sim 16$ cm, which already contain 60% of the total number of bots. These regions, about a bot-diameter distance apart from the arena wall, are probably more susceptible to bot depletion and excluded volume effects since, as we have seen previously in the density profiles, bots tend to accumulate at larger distances close to the arena wall. Both effects enhance the magnitude of fluctuations in the number of particles found in that region of the system.

8 Modelling of collective decision making

8.1 Models of opinion dynamics

Collective decision making, as we previously discussed in Chapter 1, is a process by which a collective of agents make a choice that is not directly attributable to any of the individual agents, but rather to the collective as a whole (Bose et al., 2017). Many natural and artificial systems exhibit collective decision-making and multiple disciplines study the phenomenon, including sociology, biology and physics (Dyer et al., 2009; Sasaki and Pratt, 2018). Collective decision making can be made about an infinite amount of choices. For this type of process, consensus towards an option is continuous. This can be seen in flocking dynamics when individuals have to choose a common direction of motion (Múgica et al., 2022).

But another category of collective decision making processes is one that implies finite and countable choices, and consensus reaching becomes a discrete problem. The models of collective decision making that achieve discrete consensus require a group of agents to collectively decide the best option out of n available ones (Valentini et al., 2017). In real life scenarios, examples of these types of decisions are those such as foraging patches, traveling paths or electoral candidates.

A collective decision is made when a *large majority* of agents in the group favor the same option. The threshold of what is considered a large majority is defined by the experimenter, but it must represent a cohesive collective decision with more than 50 percent of agreement amongst the individuals (Valentini et al., 2017).

Each option is characterized by attributes that make it more or less beneficial, such as size, distance or volume (for instance, in regard to a potential nesting site for honeybees). In collective decision-making models, quality and cost of each potential

option encompass these attributes. Therefore, the quality of each option represents all attributes that are relevant for the collective to achieve their objective (Garnier et al., 2009). The cost of each option is represented by the average time needed by an agent to know that that option exists. The cost impacts the decision process in an indirect form, as it is induced by the environment and it is not controlled by individual agents (Campo et al., 2010). These two properties can be configured in multiple ways in collective decision-making models. The most simplistic one is when options' qualities and costs are the same for all available options, or symmetric. This scenario deals purely with symmetry breaking by the collective selecting an option in an arbitrary way, as a result of the amplification of random fluctuations (Garnier et al., 2009). For all other scenarios, the decision-making process in the collective is shaped by the combination of qualities and costs among different options. For example, when the cost is different for each option, or asymmetric, but the qualities are symmetric, the option of minimum cost will be considered to be the best option by the collective (Schmickl et al., 2007). In the case of asymmetric qualities and symmetric costs, the maximum quality option will be chosen (Valentini et al., 2014).

Opinion dynamics models are used to understand how individuals communicate and make decisions in groups. These models involve a set of agents, each with their own opinion or state. The agents interact with each other and revise their opinions based on the opinions of others. The voter model (Holley and Liggett, 1975) is one of the simplest models used to study collective decision-making. In this model, a population of N agents are situated on the nodes of a static network, with each node representing one agent. Each agent has a binary variable $s = \pm 1$. At each time step, an agent i and a neighboring agent j are chosen randomly, and i adopts the opinion of j ($s_i = s_j$). The model assumes that each agent is influenced only by a fixed set of neighbors, with no external influences or other interactions. This revision step is repeated until the entire population of N agents reaches consensus. In finite networks, fluctuations always cause the system to eventually reach a state where the entire population holds the same opinion (Castellano et al., 2009).

The voter model has been extensively studied in many different fields, including physics (Sood and Redner, 2005), social dynamics (Castellano et al., 2009), and ecology (Martinez-Garcia et al., 2021). It is a valuable tool for exploring collective

decision-making and opinion dynamics in large populations. The voter model is particularly useful in understanding how groups of individuals can come to a consensus on a particular issue. Common metrics for measuring the voter model include the exit probability, which is the probability that all individuals will eventually adopt the same opinion, and the consensus time, which is the average time it takes for all individuals to adopt the same opinion (Castellano et al., 2009).

The majority rule model is another popular opinion dynamics model. In this model, there are N agents, each of whom can adopt one of two binary states, denoted by $s = \pm 1$. At each time step, a group of neighboring agents of odd size G is randomly chosen, and each agent in the group adopts the state of the local majority. These steps are repeated until a final consensus is reached, where all agents adopt the same opinion. This rule is a simple and intuitive way of modeling how individuals tend to conform to the opinions of those around them (Galam, 2002).

However, these models are limited in that they assume that all agents are equally likely to adopt either opinion. In reality, individuals may have different preferences, beliefs, or biases that can influence their decision-making. To address this limitation, more complex models have been developed that incorporate features such as stubbornness, partisanship, and heterogeneity (Galam, 2005; Deffuant et al., 2000). These models can help us understand how opinion diversity and polarization can arise in groups, and how these dynamics might be influenced by different factors.

8.1.1 Collective decision-making in honeybees: review of the process and models.

Insects have long been known for their fascinating behavior, and collective decision making is no exception. One particularly intriguing example of this can be found in the way honeybees choose their nest sites. This process has been the focus of many models of collective decision-making in honeybees (Britton et al., 2002; Passino and Seeley, 2006; List et al., 2009; Pais et al., 2013; Reina et al., 2017). Next, we briefly review the mechanism by which honeybees collectively decide on a new home and we further describe recent models inspired by the mechanism by which honeybees collectively choose a new home.

Honeybees collective decision-making when choosing a new nest site

Honeybees are social insects that live in large colonies, and the process of selecting a new nest site is a collective decision-making process that involves a combination of individual and group behaviors. In recent years, researchers have made significant progress in understanding the mechanisms behind this behavior, which is critical for the survival and reproduction of honeybee colonies (Seeley et al., 2012; Beekman and Oldroyd, 2018).

The process of collective decision-making in honeybees has been described primarily in regards to the species *Apis mellifera*. It begins when a colony becomes overcrowded or otherwise unsuitable. At this point, the queen and several other hundred bees will leave the colony and cluster in the surrounding area. About five percent of the swarm, the scouts, will start exploring and assessing potential nest sites based on a variety of factors, including the size and shape of the location, the quality of the nesting material, and the proximity to food sources (Seeley, 2010; Seeley et al., 2012).

After a scout bee has discovered a potential new nesting site, it will fly back to the swarm and communicate information about the prospective site to other bees through a complex dance on the surface of the swarm cluster, the *waggle dance*. Through the dance, bees recruit others to visit the site by conveying both the direction and distance to the new site, as well as its desirability. Recruited bees then visit the site and return to the cluster to perform their own dance in support of the site.

The waggle dance can consist of one to 100 or more circuits, composed by figure-eight movements made by the honeybee's body, called a return phase, and a waggling phase as it moves forward in a straight line, known as the waggle run. The duration of a waggle run ranges from 0.28 seconds to over a minute and the return phase is typically shorter than the waggle run (Seeley and Buhrman, 2001; Dyer, 2002).

The duration of the waggle dance is correlated with the honeybee's perception of the quality of the site. A longer dance corresponds to a more suitable new nest site,

while a shorter dance corresponds to a less desirable site. The intensity of the waggling and the speed of the dance can also convey information about the perceived quality of the site. For example, a more intense and rapid waggle dance can signal to other bees that the new site is of higher quality and therefore worth investigating (Seeley, 1997).

As a result, a high-quality site receives longer and more frequent advertising than a low-quality site, which leads to an overall increase in the number of bees visiting and dancing for high-quality sites, and a decrease in the number of bees doing so for low-quality sites. This highlights the importance of high-quality sites in the overall behavior and success of a bee colony. Over time the dances performed by the honeybees tend to focus on a single site. After a potential nest site has attracted enough bees, a quorum is formed. The bees that have detected the quorum will return to the swarm and produce an audible signal called piping, indicating that the decision-making process is complete. The piping serves as a message to the inactive bees in the cluster, letting them know that they should begin preparing for flight. This process helps to ensure that the entire colony is in sync and ready to move to their new home (Beekman and Oldroyd, 2018).

Models of collective-decision making in honeybees

Models of nest site selection in honeybees often consider factors such as the quality of the different nest sites, the number of scouts who have visited each site, and the persistence of the scouts in advertising their preferred sites.

Through the study of these models, researchers can gain valuable insights into the underlying mechanisms of honeybee decision-making and the factors that contribute to the process of achieving consensus. Moreover, these models have the potential to inform our understanding of group decision-making in other animals, including humans. For instance, as shown in Pais et al., 2013, models of honeybee swarm decision-making can be applied to various levels of biological complexity, from individual brains to intracellular decision-making circuits, and even be utilized to inform the design of artificial, decentralized decision-making systems.

The model deals with binary-choice decision-making and involves agents that can be in one of three states: uncommitted to either alternative ($s = 0$), committed

to state 1 ($s = 1$), or committed to alternative state 2 ($s = 2$). Each agent interacts stochastically with every other agent in several ways, each interaction having a particular probability per time step.

In this model, there are four types of interactions: commitment, abandonment, recruitment, and direct switching. An uncommitted agent ($s = 0$) may spontaneously commit to either state with probabilities γ_1 and γ_2. A committed agent to either state can spontaneously abandon its commitment and return to an uncommitted state with probabilities α_1 and α_2. A committed agent to either state may recruit an uncommitted agent to the state that the former is promoting with probabilities ρ_1 and ρ_2. Alternatively, a committed agent to either state can switch another agent's state, who is already committed to the alternative state, such as in the recruitment interactions, but with probabilities δ_1 and δ_2. The total number of agents is fixed as $N = N_1 + N_2 + N_0$, and all rates $\gamma_j, \alpha_j, \rho_j, \delta_j$ are greater than 0.

In this model, it has been found that when faced with a decision between two equal states, the system inevitably reaches a stable deadlock, with an equal number of agents committed to each alternative. As one site gains a majority of agents, the rate of switching from that site to the other increases, ultimately forcing the system back into a state of equal commitment. However, a stable deadlock is suboptimal as it fails to reach a consensus. To address this issue, a modification to the model has been proposed that replaces the **direct switching** interaction with a **discriminate stop signal** interaction. This change results in a significant qualitative difference, allowing the system to overcome deadlocks when faced with equal alternatives.

Under the **discriminate stop signal** interaction, an agent committed to a state can actively inhibit the commitment of an agent committed to the other state, convincing the latter to lose its commitment and become uncommitted with probabilities σ_1 and σ_2. This modification enables the system to move away from the deadlocked state and towards a consensus.

Reina et al., 2017 extended the original model from the study of binary choices to the study of the best-of-N options, and by introducing a new control parameter r. This parameter represents the ratio between the individuals' interaction and the spontaneous transitions (i.e. the probability that a scout bee chooses a nest site at random, independent of the information gathered from other bees). The definition of

r has a simple and natural biological interpretation, as it represents the propensity of scout bees to deliver signals to others, relative to the rate of spontaneous transitions. The larger the value of r, the greater the effect of social influence on the decision-making process.

The introduction of the new parameter r in the model highlighted the potentially high adaptability of honeybees' decision-making process to changes in the environment. The model demonstrated that honeybees become committed to the higher quality option when the value of one option exceeds that of the other, and the commitment to the low quality option decreases until it is abandoned altogether. Moreover, the model showed that when bees face equal quality options, increasing the signaling behavior has the potential to break decision deadlocks. Thus, it might be prudent to increase the value of r over time. However, excessive signaling can reduce decision accuracy when deciding among options of varying quality. The study showed that while the level of signaling required to break deadlocks between N equal options increases quadratically with N, the level of signaling that optimizes the discriminatory ability of the swarm in best-of-N scenarios increases only linearly. Thus, optimizing the best-of-N decisions seems to conflict with optimizing equal alternatives scenarios. However, in natural environments, the probability of encountering N equal-quality nest options decreases rapidly with N.

The results from this model suggest that honeybees may use a combination of strategies in different contexts to optimize their collective decision-making process. This observation was initially reported by List et al., 2009 in their proposed model of house-hunting honeybees. This model is the first agent-based model that integrates the interaction between independent individual decisions and imitation to capture the decision-making of honeybees in the process to select a new site to build their nest. In the following section, we elaborate on the key features of this model.

8.2 Agent-based model of nest-site selection by honeybees

The nest-site choice model proposed by (List et al., 2009) consists of a swarm ranging from 1 to N scout bees that collectively decide towards one of the potential nest sites, labeled $1, 2, 3, ..., k$ where each site j has an intrinsic quality $q_j \geq 0$ that reflects on the

time a bee is advertising site j. In each discrete time step Δt (in this case, $\Delta t = 1$), the status of bees is updated in parallel. Each bee's state at $t + 1$ depends both on the individual and the other bees state at time t.

At time t, a bee can either be dancing (i.e. promoting) for one of the k sites, or not dancing for any site, which could represent that they are still searching for a site, observing other bees dance or resting. Formally, the vector $x_{i,t} = (s_{i,t}, d_{i,t})$ represents the state of bee i at time t, where $s_{i,t}$ refers to the site it is dancing for, with $s_{i,t} = 0$ meaning that she is not dancing for any site, and $d_{i,t}$ the remaining duration of the dance for bee i.

The model starts at time $t = 1$ where none of the bees are committed to a site, that is $x_{i,1} = (0, 0)$, although, in the steady state, the model is insensitive to initial conditions and would yield similar results even if it started with different proportions of committed and uncommitted bees.

Bees that are not dancing at time t have a probability of starting to dance for site j at time $t + 1$, denoted $p_{j,t+1}$. This adds up to 1 including the probability of keeping uncommitted to any site at $t + 1$, $p_{0,t+1}$. $p_{j,t+1}$ depends on the likelihood for a bee i to find site j independently of another bee promoting the site, and in proportion of bees dancing for site j at time t. It is calculated as follows

$$p_{j,t+1} = (1 - \lambda)\pi_j + \lambda f_{j,t} \tag{8.1}$$

Here, λ refers to the interdependence of each bee to the rest of the collective, π_j to the *a priori* probability of discovering site j independently, and $f_{j,t}$ represents the proportion of bees dancing for site j at time t. The term λ ranges between 0 and 1. This determines how much bees rely on each other to decide to dance for a site. If $\lambda = 0$, the probability of finding site j becomes the *a priori* discovery probability regardless of the proportion of bees dancing for it. If $\lambda \to 1$ the probability of committing to site j at $t + 1$ depends almost entirely on the proportion of bees dancing for it at time t, $f_{j,t}$. The particular case with $\lambda = 1$ is ill-defined in this model, since new self-discovery information would not be introduced in the swarm. Thus, at $t + 1$, each non-dancing bee will start dancing for one of k sites with probability $p_{j,t+1}$, and will continue uncommitted to a site with probability $p_{0,t+1}$. The original

model features a dance behavior in which a bee's dance duration for a particular site is fixed either through an independent assessment of the site's quality, with exactly q_j steps, or by mimicking others, which results in a generic dance time. For simplicity, in our modified version of the model if a bee starts dancing for a site they will dance for exactly q_j time steps; therefore, the duration of the dance is fixed by the site quality.

Bees that are already dancing for a site at time t will continue dancing for it at $t+1$ if $d_{i,t} > 1$, reducing $d_{i,t}$ by one timestep, and will stop dancing if $d_{i,t} = 0$,

$$x_{i,t+1} = \begin{cases} (s_{i,t}, d_{i,t} - 1) & \text{if} \quad d_{i,t} > 1 \\ (0,0) & \text{if} \quad d_{i,t} = 1 \end{cases} \tag{8.2}$$

8.2.1 Analytical description of a further simplified model

Following the publication of the original model, Galla, 2010 proposed a simplification which replicates the same features as the original system, and whose stationary state can be found analytically in the mean-field approximation. The model reduces the state variables from $(s_{i,t}, d_{i,t})$ to only $(s_{i,t})$, which can then be treated as a Markov chain.

The simplification implies substituting the dance time by a stop-dancing rate. In List et al., 2009 model, once a bee starts dancing the end of the dance occurs at a deterministic moment in time. In the reduced model proposed by Galla, 2010 this is replaced by a random process under which a bee can stop dancing with a probability r_j and continues dancing with a probability $(1 - r_j)$. The stopping rate is inversely proportional to the site quality $r_j \propto q_j^{-1}$. The probability that a dance for a given site j lasts precisely l time steps is given by $(1 - r_j)^{l-1} r_j$, which gives a geometric distribution of mean $1/r_j$. The modification of the original algorithm occurs at each state update from t to $t+1$ for bees that are dancing at t. Non-dancing bees compute state probabilities as in the original model. In the simplification, if bee i is dancing for site j at time t (i.e. $s_i(t) = j > 0$), then with probability r_j set $s_i(t+1) = 0$ (bee stops dancing), and with probability $1 - r_j$ set $s_i(t+1) = j$ (bee keeps dancing for site j).

To specify the stopping rates r_j, Galla uses a definition that accounts for the degree of bees' independent quality assessment of a site. As previosly mentioned, our modified List et al., 2009 model does not include this parameter, but rather a fixed dance duration $d_j(t) = q_j$. For this, our definition of stopping rates is only determined by site quality:

$$r_j = \frac{q_0}{q_j} \tag{8.3}$$

where q_0 is a prefactor that defines the time scale of the model (typically set to 1), and with its upper bound at the lowest site quality, which ensures that $0 < r_j \leq 1$. The mean duration of the dance for site j, $1/r_j$ is proportional to the site's quality, as it is originally defined in the List et al., 2009 model.

The simplified stochastic model defines a one-dimensional Markov process that can be analyzed in terms of the master equation. First, the state of the system is defined as the number of bees dancing for each site $n_j(t)$, including $j = 0$, i.e. bees not dancing at time t. Thus, one has $\sum_{j=0}^{k} n_j(t) = N$ for all times t. For convenience, Galla, 2010 introduces the shorthand vector notation $\mathbf{n}(t) = (n_0(t), n_1(t), ..., n_k(t))$. The state of the system at time t is therefore fully determined by $\mathbf{n}(t)$. A master equation can be then proposed for the probability of finding the system in state $\mathbf{n}$:

$$\begin{aligned}\frac{d}{dt}P_{\mathbf{n}} = \sum_{j=1}^{k} P_{\mathbf{n}-\mathbf{e_j}+\mathbf{e_0}} T_j^+(\mathbf{n}-\mathbf{e_j}+\mathbf{e_0}) - P_{\mathbf{n}} \sum_{j=1}^{k} T_j^+(\mathbf{n}) \\ + \sum_{j=1}^{k} P_{\mathbf{n}+\mathbf{e_j}-\mathbf{e_0}} T_j^-(\mathbf{n}+\mathbf{e_j}-\mathbf{e_0}) - P_{\mathbf{n}} \sum_{j=1}^{k} T_j^-(\mathbf{n})\end{aligned} \tag{8.4}$$

for the probability $P_{\mathbf{n}}(t)$ of finding the system in state $\mathbf{n}$ at time t. $T_j^+(\mathbf{n})$ stands for the probability that an inactive bee starts dancing for site j at time t, given the current state $\mathbf{n}$ of the system. Likewise, $T_j^-(\mathbf{n})$ is the probability for a bee dancing for site j to stop dancing in the next time step. $\mathbf{e_j}$, for $j = 0, 1, ..., k$, are the unit vectors representing each site, i.e. $(e_j)_m = \delta_{j,m}$ for $j, m \in 0, 1, ..., k$ The transition rates are defined as:

$$T_j^+(\mathbf{n}) = n_0 \left[(1-\lambda)\pi_j + \lambda \frac{n_j}{N} \right] \tag{8.5}$$

and

$$T_j^-(\mathbf{n}) = n_j \left[\frac{q_0}{q_j} \right] \tag{8.6}$$

The transition rate $T_j^+(\mathbf{n})$ is proportional to the number of non-dancing bees n_0 and, similarly, T_j^- is proportional to the number of bees dancing for site j, n_j.

Finally, to derive a set of deterministic ordinary differential equations in the limit of an infinite system size, $N \to \infty$, continuous frequency variables $f_j = n_j/N$ are introduced. For instance, considering

$$\langle \mathbf{n}(t) \rangle = \sum_n \mathbf{n} P_{\mathbf{n}}(t) \tag{8.7}$$

and working through the expression of its time derivative, one arrives at

$$\begin{aligned} \frac{d\langle \mathbf{n} \rangle}{dt} &= \sum_{j=1}^{k} \left[(\mathbf{e_j} - \mathbf{e_0}) T_j^+(\langle \mathbf{n} \rangle) \right] \\ &+ \sum_{j=1}^{k} \left[(\mathbf{e_0} - \mathbf{e_j}) T_j^-(\langle \mathbf{n} \rangle) \right] \end{aligned} \tag{8.8}$$

where a deterministic approximation has been used to write $\langle T_j^{\pm}(\mathbf{n}) \rangle = T_j^{\pm}(\langle \mathbf{n} \rangle)$. Using this expression for each site, one finds

$$\dot{f}_j(t) = (1 - \rho(t))[(1-\lambda)\pi_j + \lambda f_j] - f_j r_j \tag{8.9}$$

where $\rho(t) = \sum_{j=1}^{k} f_j(t)$, and the fraction of bees not dancing at time t is given by $f_0(t) = 1 - \rho(t)$. Equation 8.9 can be integrated numerically for any fixed choice of parameters. Nevertheless, an expression for the stationary points of this equation can be found as the solution of k coupled quadratic equations,

$$r_j f_j^* = \left(1 - \sum_{m=1}^{k} f_m^* \right) [(1-\lambda)\pi_j + \lambda f_j^*] \tag{8.10}$$

This equation can be rewritten in the form,Galla, 2010:

$$f_j^* = \left[\frac{r_j}{1-\rho^*} - \lambda \right]^{-1} (1-\lambda)\pi_j \quad j = 1, ..., k \tag{8.11}$$

The expressions for the different f_j^*, for $j = 1, ..., k$, can be combined to first find a closed equation for the stationary value of f_0,

$$f_0^* = 1 - (1 - \lambda) \sum_{j=1}^{k} \left[\frac{r_j}{f_0^*} - \lambda \right]^{-1} \pi_j \tag{8.12}$$

Eq.8.12 has $k + 1$ roots (which can be found by solving the equation numerically or by rearranging it as a $(k + 1)$th degree polynomial in f_0), but not each one leads to physical solutions, as some of them result in $f_0^* > 1$. From the remaining valid solutions $f_0^* \leq 0$ only one will lead to a valid and linearly stable solution for the rest of dance frequencies, $f_j^* \leq 1$ (once f_0^* is introduced into Eq. 8.11). Stochastic simulations and the integration of Eq. 8.9 confirm the stability of this result. In the particular case of $k = 2$ sites one always find one non-physical f_0^* solution ($f_0^* > 1$) and from the two remaining only the lowest one leads to valid solutions for the rest of f_j^*.

The limit $\lambda = 0$ in Eq. 8.12 leads to the simpler solution,

$$f_0^* = \frac{1}{1 + \sum_{m=1}^{k} \pi_m q_m}, \tag{8.13}$$

that we can use to compute the result for the rest of the dancing frequencies. Using Eq. 8.11, we obtain

$$f_j^* = \frac{\pi_j q_j}{1 + \sum_{j=m}^{k} \pi_m q_m} \tag{8.14}$$

Another special case is when $\lambda \to 1$. As mentioned, $\lambda = 1$ is an ill-defined case where the dynamics of the system remains the same for any value the discovery probabilities π_j's. Nevertheless, it is worth looking at the limit when this parameter approaches the value 1. Due to the extreme reliance on interdependence, one expects that the site with a greater quality will be finally dominating the whole system, leaving no agents committed to the other sites and only a small quantity of uncommitted agents (remember that in this model, bees can always finish their dance, even if they dance for the best-quality site, and return to the uncommitted state for a few iterations). Then, assuming that $q_k > q_{k-1} > ... > q_1$ we can impose that

$f_1^* = ... = f_{k-1}^* = 0$, and using 8.10 we find the following stationary solution:

$$f_0^* = r_k, \quad f_1^* = ... = f_{k-1}^* = 0, \quad f_k^* = 1 - r_k \tag{8.15}$$

This result can be validated by simulating the system, or by solving the equations at high values of λ. It shall be noticed, though, that the opposite solutions where one of the lower quality sites is imposed, for instance site $k-1$, i.e. $f_0^* = r_{k-1}$, $f_1^* = ... = f_{k-2}^* = 0$, $f_{k-1}^* = 1 - r_{k-1}$, $f_k^* = 0$, are mathematically viable. To confirm the validity of the stationary solutions 8.15, we can also perform a linear stability analysis (LSA) of the possible mathematical solutions. Without delving into too much detail, we find that performing a LSA on the first solution (the best-quality site k imposes itself in the system) results in all negative eigenvalues, i.e. it is an stable solution. Instead, when performing a LSA to any other alternative fix point that imposes any of the $j \neq k$ less-quality sites, we find that there are at least $k - j$ positive eigenvalues, meaning that these solutions are actually saddle points - which are never reached by numerical simulation.

8.2.2 Mean-field parametric exploration of the nest-site selection model

First, we confirm that stationary states for the original model by List et al., 2009 can be obtained by stochastic simulations of the simplified model, by the numerical integration of Eqs. 8.9, and from the combination of the numerical solution in Eq. 8.11 by first solving Eq. 8.12. Moreover, we explore the model's parametric space to check for consistency, and to gain a deeper understanding of the effect that each parameter has on consensus reaching in mean-field approximations.

For the parameter exploration, we fix the number of agents to $N = 100$ and the number of potential nest sites to $k = 2$ for the simplest nest selection problem. Site quality values q_j, which are directly correlated with dance duration, were chosen based on real-life honeybee scenarios (Grüter and Farina, 2009). Of the two potential nest sites, site 1 represents low and medium quality sites, with quality values q_1 ranging between 3 and 9, and site 2 represents the high quality site, with a quality value $q_2 = 10$. We explore a wide range of *a priori* discovery probabilities for both

sites$\pi_{1,2} \in [0.01, 0.5]$, to understand the effect of different likelihoods of bees becoming scouts, and independently finding a site. The exploration of the parameter space defined by discovery probability values is simultaneously carried out for varying values of the interdependence parameter λ (the likelihood of a bee becoming a scout could be seen as $(1-\lambda)\pi_j$ in the p_j equation). To analyze the collective decision making, we measured both dancing frequencies f_j, and a strong majority definition of consensus,

$$Q = f_2 - 2f_1 \tag{8.16}$$

This implies that there must be twice as many bees dancing for the high quality site than for the low quality site, where $Q > 0$ represents a strong consensus towards site 2, and $Q < 0$ indicates that less than $2f_1$ are dancing for site 2. Here we apply this strong consensus definition instead of a simple one (i.e. $Q_s = f_2 - f_1$), as it ensures that the fraction of bees dancing for site 2 represent a *large majority*, i.e. a majority over a factor of 2/3s if there were no uncommitted bees in the system.

First, we observe how dance frequencies f_j evolved in time for stochastic simulations and from the numerical integration of Eqs. 8.9. Figure 8.1 shows average values of the frequencies f_0, f_1 and f_2 as a function of time for the stochastic simulations and from the numerical integration of the deterministic mean-field equations for a fixed set of parameters. Simulations of the stochastic model show oscillatory fluctuations of all dance frequencies and their convergence towards a such fluctuating state. The numerical integration produces smooth evolution curves that does not accurately characterize the transitory state, but describe properly the average stationary value obtained by the stochastic model.

We then explore the effect of *a priori* discovery probabilities in both f_2 and consensus reaching for different levels of interdependence. For this, we match $\pi_1 = \pi_2$ and scan the parameters $(\pi_{1,2}, \lambda)$ space. We fix the values of the site qualities close to each other, i.e. $q_1 = 7$ and $q_2 = 10$. This was done to have clear observations of the transition from non-consensus to consensus reaching. We then plotted the averaged stationary values of f_2 and Q calculated over 100 realizations (Figure 8.2 A and B). The color charts illustrate the corresponding values of these two magnitudes throughout the parameter space. The orange line on the right plot corresponds to

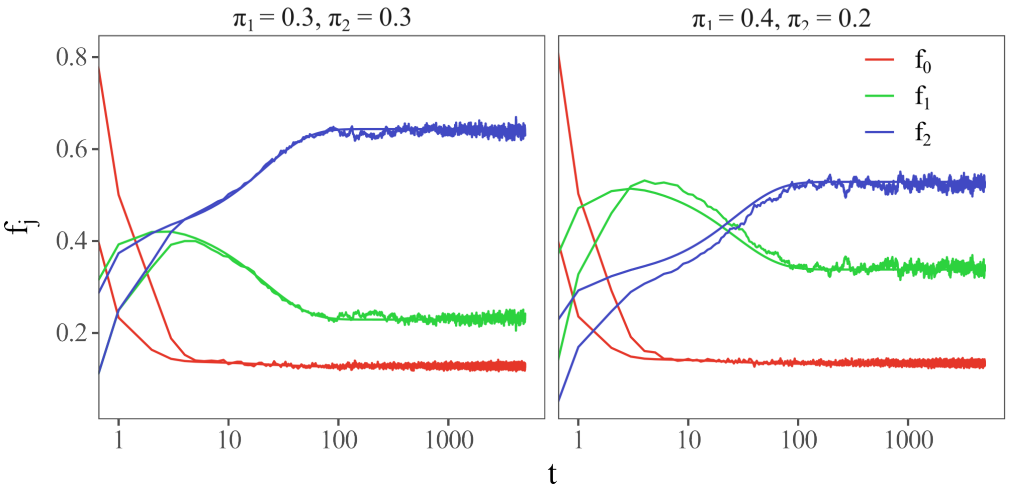


Figure 8.1: Frequencies f_0 (red), f_1 (green) and f_2 (blue) as a function of time obtained from numerical simulations of the stochastic model, and from the numerical integration of the mean-field deterministic equations (smooth superimposed curves in the same color). Left: Symmetric discovery scenario with probabilities $\pi_1 = \pi_2 = 0.3$. Right: Asymmetric discovery scenario with probabilities $\pi_1 = 0.4$ and $\pi_2 = 0.2$. Simulations were averaged over 100 realizations. Other parameters are: $N = 100$, $q_1 = 7$, $q_2 = 10$, and $\lambda = 0.6$.

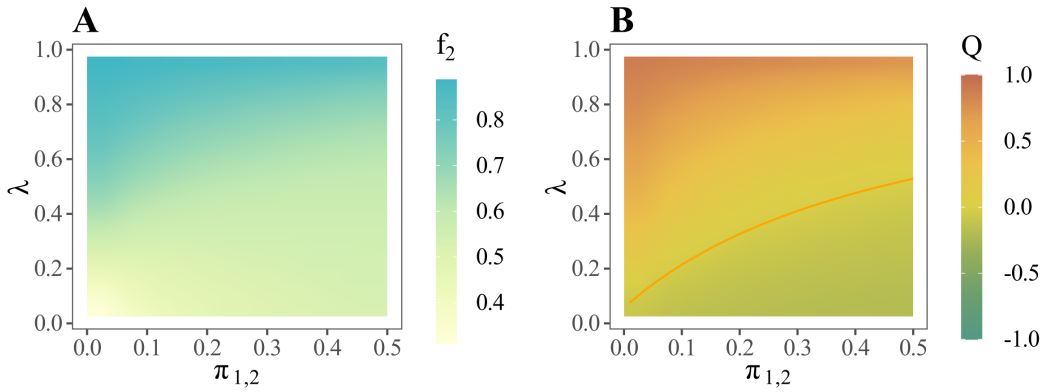


Figure 8.2: **A**: Stationary values of f_2 in the parameter $(\pi_{1,2}, \lambda)$ space obtained from numerical simulations of the stochastic model. **B**: Stationary value of Q in the parameter $(\pi_{1,2}, \lambda)$ space obtained from numerical simulations of the stochastic model. The orange line corresponds to the theoretical crossover line where $Q = 0$. Here $N = 100$, $\pi_1 = \pi_2$, $q_1 = 7$, and $q_2 = 10$.

the theoretical crossover line where the strong consensus parameter $Q = 0$.

The proportion of agents dancing for the high quality site f_2 increases with the interdependence parameter. Independent discoveries for both sites introduce more agents dancing for site 1, which decreases the ratio of bots dancing for site 2 as the value of $\pi_1 = \pi_2$ increases. This translates in strong consensus not being achieved (i.e. $Q < 0$) when interdependence is very low, or lower than 0.5. After a $\pi_{1,2}$-dependent λ threshold, the system crosses over to a consensus situation, $Q > 0$, and becomes stronger as interdependence increases. The region of non-consensus

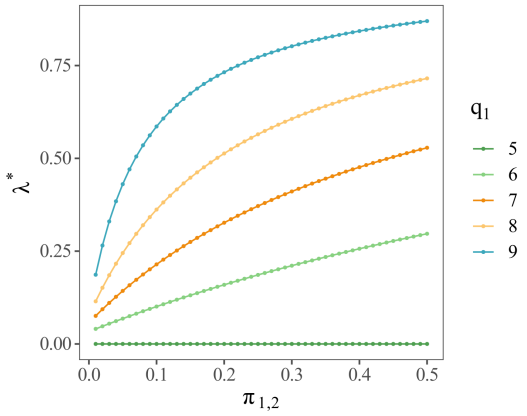


Figure 8.3: Theoretical curve representing the behavior of the consensus crossover threshold value λ^*, i.e. $Q(\lambda^*) = 0$, as a function of $\pi_1 = \pi_2$. Colors represent different values of the low-quality site q_1. Other parameters are $N = 100$, $q_2 = 10$.

grows as $\pi_{1,2}$ grows, narrowing the range of interdependence for which consensus is reached. This is a consequence of bees having a higher probability of independently discovering the less quality site, and therefore having more bees dancing for it. Nevertheless, higher levels of interdependence increase the proportion of bees dancing for the high quality site, which is benefited by longer dance times. Figure 8.2 also shows the theoretically predicted crossover curve from $Q = 0$. This curve is obtained by iteratively solving Eq. 8.12 for f_0, then computing the corresponding $f_{1,2}$ and finally Q, changing λ until Q converges to 0. The stochastic simulations accurately align with the null strong consensus curve.

In Figure 8.3, we represent the theoretical curve representing the behavior of the consensus crossover threshold value λ^*, i.e. the interdependence parameter λ^* for which $Q(\lambda^*) = 0$, as a function of $\pi_1 = \pi_2$. Different curves, from cyan blue to dark green, correspond to increasing values of q_1 and a fixed value of $q_2 = 10$. From the figure, one can appreciate the dependence on quality difference between the two sites. On one hand, when there is a site that is perceived as of higher quality than its competing site, in this case $q_1 <= 5, q_2 = 10$, the swarm discriminates the high quality site for all λ values. This occurs even if the sites' independent discovery increases the number of bees dancing for the low quality site. On the other hand, when q_1 increases and gets closer to q_2, the competition between sites intensifies, and consequently a higher λ is needed to counteract the effect of *a priori* discovery probabilities.

To further explore the statistical behavior of the model, we studied the case of asymmetric *a priori* discovery probabilities by plotting average values of f_2 and Q for different π_1 and π_2 from 0 to 0.5, now with a fixed value of λ. Figure 8.4 shows stationary values at three contrasting values of interdependence $\lambda = 0.3, 0.6, 0.9$, and the same site qualities $q_1 = 7$ and $q_2 = 10$ as for the symmetric case. In all plots, the diagonal starting on the bottom-left corner and ending on the upper-right corner represents the symmetric $\pi_1 = \pi_2$ scenario and, thus, divides situations where the high quality site is easier to find independently (upper triangle where $\pi_1 < \pi_2$) from those where the low quality site has more probability to be found (lower triangle with $\pi_1 > \pi_2$). When the high quality site is easier to discover than the low quality site, we expect the dance frequency for site 2 to be high and the system to reach a good consensus for the high quality site ($Q > 0$). This is indeed the case when $\pi_1 << \pi_2$. Nevertheless, even if discovering the best site is easier, if both *a priori* discovery probabilities are high, the ratio of agents dancing for the high quality site is lower and consensus reaching becomes more challenging as new bees dancing for the worst site are constantly introduced in the system. This is particularly applicable for low interdependence values (e.g. $\lambda = 0.3$), where consensus is reached for very low π_1 but if both sites are easier to discover, a strong majority cannot be achieved. When interdependence is intermediate (e.g. $\lambda = 0.6$) consensus is achieved even for relatively similar discovery probabilities, but when $\pi_1 >> \pi_2$ the frequency of bees dancing for site 1, f_1, prevents the high quality site from permeating the swarm. In Figure 8.4, we can also see a region for small π_2 where strong consensus is achieved, even if $\pi_1 > \pi_2$. In these situations, strong consensus can still be achieved through information exchange, as the dance times for site 2 last longer, and, are therefore broadcasted for a longer period of time to other bees. Exceeding a λ-dependent π_1 threshold, the dance duration for site 2 cannot longer surpass the frequent independent discovery and advertising of site 1. For this parameter combination, high interdependence (e.g. $\lambda = 0.9$) ensures consensus in all conditions, even when $\pi_1 >> \pi_2$.

As seen in Figure 8.1, a consequence of the agents' constant state update is that stationary values in numerical simulations are fluctuating, and thus distributed around

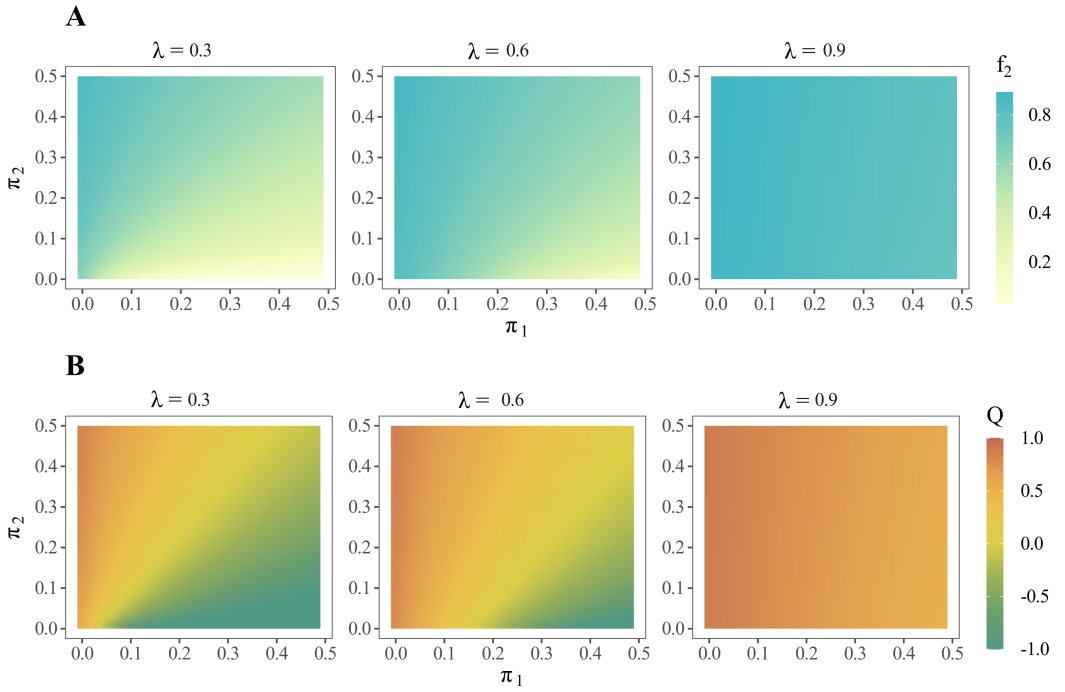


Figure 8.4: **A**: Stationary values of f_2 and **B**: Stationary values of Q in the (π_1, π_2)-parameter space. From left to right, we represent stochastic simulation results for increasing values of the interdependence parameter $\lambda = 0.3, 0.6$ and 0.9. Here $N = 100$, $q_1 = 7$, and $q_2 = 10$. Stationary values were averaged over 100 realizations.

central values. We also explored the associated variability in the form of the standard deviation (σ) of f_2 over the $(\pi_1 = \pi_2, \lambda)$-parameters space in the symmetric discovery scenario. We also explored finite-size effects in the behavior of those fluctuations by considering three different group sizes with $N = 35, 100, 500$ bees. Here we also include the small system size of $N = 35$ bees because we wanted to characterize the behavior of a system size comparable to our experimental $35 - kilobot$ swarm. Experimental results will be discussed in Chapter 9.

In Figure 8.5, one can clearly see that regions with high variability concentrate around and above the transition to consensus. Therefore, although average stationary consensus towards the high quality site has been reached, i.e. $Q > 0$, the errors represented by the value of σ indicate that in a significant number of realizations consensus has not been reached. Comparing σ values among different system sizes, we can also conclude that fluctuations are mainly due to finite size effects of the system as they decrease when N increases, contrary to what would be expected if we were facing a second order phase transition. The increase in consensus accuracy for larger groups is congruent with the *wisdom of the crowds* effect that relies on large

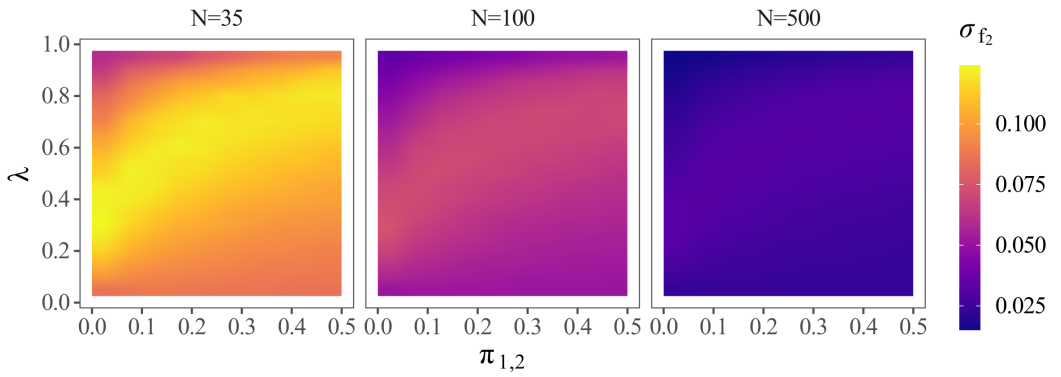


Figure 8.5: Standard deviation, σ_Q, of the strong consensus parameter in the $(\pi_{1,2}, \lambda)$-parameter space. From left to right, we represent increasing values of $N = 35, 100, 500$. Site qualities were fixed at $q_1 = 7$, and $q_2 = 10$. Standard deviation calculated over 100 realizations.

group sizes to cancel the inherent noise of individual opinions to make a smarter collective decision.

8.3 Closing remarks

Models of collective decision-making in honeybees have provided valuable insights into the underlying mechanisms of honeybee decision-making and the factors that contribute to the process of achieving consensus. From the models we have learn that honeybees may use a combination of strategies in different contexts to optimize their collective decision-making process and to break possible decision deadlocks. Moreover, these models have the potential to inform our understanding of group decision-making in other animals, including humans. Further research in this area may lead to new ways of designing artificial, decentralized decision-making systems.

By means of numerical mean-field simulations we have categorized the behavior of a decentralized decision making model that seeks to shed light on honeybees' mechanisms to choose the best option when looking for a new nest site. The model integrates both the independent discovery of potential nest sites and inter-individual communication to achieve consensus towards the best site. Sites qualities are then expressed in dance duration: the higher the quality, the longer an individual promotes it. The difference in dance times is a driving indicator for breaking the symmetry of the system. This is because the high quality sites receive more "advertisement",

which in turn generates more dancing activity for the site. The interdependence parameter helps the swarm to make a collective decision toward a higher quality site. When each site is equally likely to be discovered, a high interdependence is beneficial in order to counteract the continuous surge of individuals dancing for the low quality site, specially when the probabilities of sites being discovered are high.

When sites have asymmetric *a priori* discovery probabilities, if the high quality site has a higher probability to be found, even without any communication the swarm will reach consensus towards it. The scenario where communication becomes crucial is when the low quality site has a higher probability to be discovered than the high quality site, i.e. $\pi_1 > \pi_2$. The level of communication needed to achieve consensus towards the high quality site increases with π_1 and with its distance with respect to π_2. In this model, independence acts as a random noise in the process that hampers the ability of the system to reach consensus, specially when sites are easy to discover.

9 Honeybees nest-site selection model in kilobots

The original List et al., 2009 model seeks to unravel the underlying behavior of individual honeybees from which consensus can emerge when they are choosing a new nest site. Nonetheless, the model uses a mean field approach where agents have access to all the information of the system at each time step. This is an unrealistic assumption. In general, living organisms have sensor and cognitive limitations regarding the distance and number of individuals they can pay attention to. In particular, honeybees interact with only a number of neighboring bees when they are dancing or when they are watching a waggle dance (Sumpter, 2006; Judd, 1995).

To explore the feasibility of the model as a possible mechanism for real honeybees nest-site selection, we analyze the effects that more realistic interactions might have in consensus reaching. We do this by adding space and locality to the original model proposed by List et al., 2009, and by applying the model to study consensus formation in a physical system. To do so, we have chosen kilobots as our study subjects as they are designed as swarm robots (Rubenstein et al., 2012; Valentini et al., 2016; Gauci et al., 2018) . As discussed previously in Chapter 7, this is a system that consists of multiple robots that can perform collective tasks through their capability of exchanging information on relatively short-range distances, i.e. within a radius of approximately 7 cm, while moving through space.

The study of insects via robotics has been widely applied as it has several advantages. Particularly, the inherent behavioral variations of the robots, and that they interact with environmental stimuli, which introduces real noise within the model. Therefore, the introduction of artificial noise to simulate environmental conditions is not required, which increases the robustness of the model proposed mechanisms. Moreover, if the robots manage to *solve* the proposed problem of the model, then the

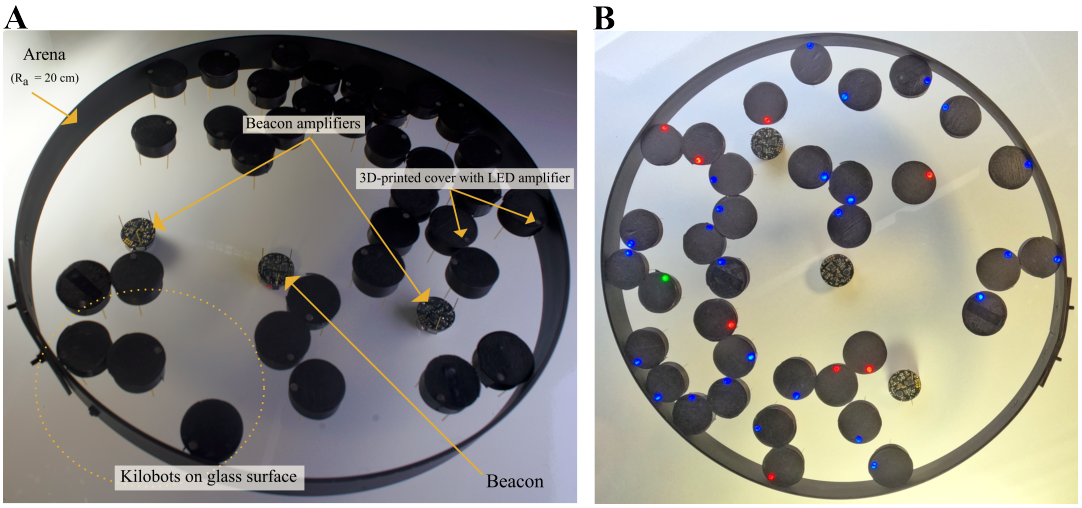


Figure 9.1: A: Kilobots experimental setup. **B**: Kilobots performing the modified List et al., 2009 model. Each kilobot is covered by a custom 3D-printed case, and some are dancing for the low quality site (green), others for the high quality site (blue) or not dancing for any site (red).

noise they were subjected to is not enough to preclude the solution (Sharkey, 2007).

9.1 Experimental Setup

In this case, the experimental setup consisted of a whiteboard melamine base and a glass surface held 15 cm from the base, where kilobots performed their motion. In the melamine base, we placed a central kilobot, turned upside-down, that functioned as a beacon to mark the clock to all other kilobots sending messages via infrared signals. Two additional kilobots were also placed on the base to amplify the signal of the beacon as shown in Figure 9.1.

In the absence of the beacons, the kilobot swarm exhibits synchronization problems within the long time runs. The requirement for synchronization originates when processes, such as the ones investigated in this chapter, need to be executed concurrently. Thus, the main purpose of enabling the synchronization of the kilobots internal clock is the coordination of the process iterations.

Our experiment consisted of a group of 35 kilobots moving as persistent random walkers (PRW) in a circular arena of radius, $R = 20$ cm as described in Chapter 7. After confirming that the kilobots were able to receive and transmit messages, and had sifficient battery-life, we started our experiments to observe consensus-reaching.

For the kilobots to gather information from their surroundings, they need to have sufficient time to sense their local environment, and also enough time for their set of near-neighbors to change. With this in mind, and with consideration of the kilobot motion and message-passing characteristics explored previously in Chapter 7 (see, for instance, Figures 7.10 and 7.6), we established a time step of $\Delta t = 800$ kilobots' loop iterations, which translates into approximately 8.24 seconds. During this time, our group of kilobots are constantly transmitting and receiving messages. Moreover, for each time step Δt signaled by the beacons, kilobots were either dancing for site 1 (the low quality site), for site 2 (the high quality site) or not dancing for any site. We utilize the kilobot's built-in led lights to identify each bot state: red for non-dancing bots, green for bots dancing for site 1, and blue for bots dancing for site 2, as shown in Figure 9.1 B. Kilobots that were not dancing for any site did not move while other bots promoted the site they were dancing for by performing a persistent random walk with discrete angles, as described in Chapter 7.

To identify the state of the dancing bots, we isolate the kilobots' led lights from other light sources. To achieve this, we recorded kilobots in absolute darkness and built custom made 3D-printed black casings to cover the kilobot body except the led light, which prevented the led from reflecting on the metal components of the kilobot as shown in Figures 9.1 A and B.

Recordings were made with a video camera Sony Alpha a7rii with a spatial resolution of $1920x1080$ pixels and a temporal resolution of 25 frames per second. We recorded the kilobots' decision making process from a top-down view for 30-minute intervals for each realization of all the experimental conditions considered. We extracted one image frame at the second half of every Δt cycle of 8.2 seconds up to 210 images, which we identified as a sufficient number of time steps for the system to reach a stationary state (see for instance Figure 9.5A).

9.1.1 Kilocounter: A color tracking software

To automatically count the number of kilobots dancing for each site at each time step, a color tracking software, kilocounter, was developed. Kilocounter is designed to identify and count the number of bright or dark colored blobs in an image. The software utilizes the scikit-image image processing library, which is implemented in

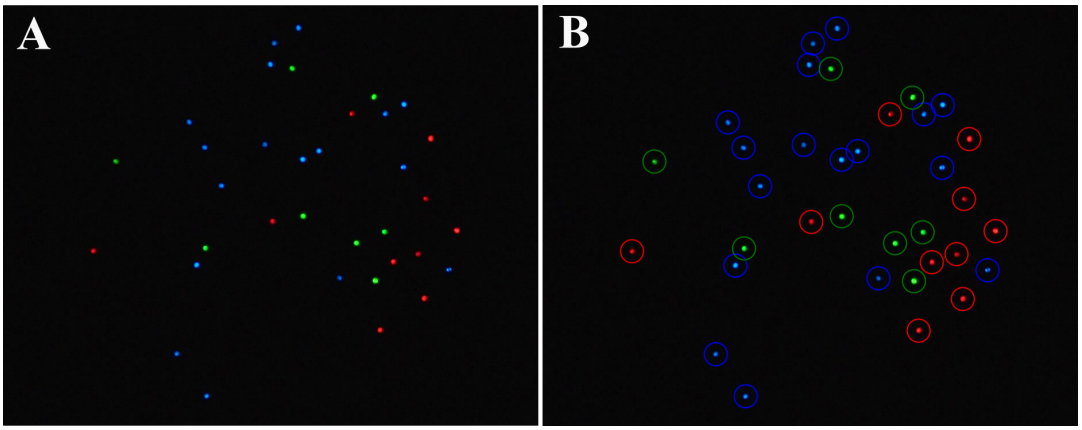


Figure 9.2: **A**: A single frame showcasing a group of 35 kilobots, each identified by their colored light, either dancing for site one (green), site two (blue), or not dancing for any site (red). The frame was recorded in a dark setting. **B**: The same frame as in A, processed using the kilocounter software, which has identified 8 bots with green lights, 17 bots with blue lights, and 10 bots with red lights.

the programming language Python. During the tracking of our experimental recordings, the number of kilobots present in the arena is known, and this information is provided to the Kilocounter software as an input parameter. This allows the software to optimize the detection process, as it can identify the exact number of kilobots to track.

To speed up the image detection process, we use a Multidimensional Gaussian filter in combination with the peak_local_max function to detect blobs with specific colors within the image. The Gaussian filter is a widely used image filtering technique that consists of a low-pass filter that blurs the image by convolving it with a Gaussian kernel. This has the effect of smoothing the image and reducing noise, which makes it easier to identify the blobs (George et al., 2018).

The Gaussian filter is controlled by a parameter σ that modifies the width of the kernel function used in the filter; larger values of σ produce a wider kernel function that blurs over a wider radius. Choosing an appropriate value for σ is important to ensure that the filter is able to cover all the objects in the image, but not so large that it overlaps neighboring objects. However, there is no formula to determine the optimal value of σ, as it depends on various image factors, including the resolution of the image and the size of the objects being detected (in pixels; Walt et al., 2014). In our study, we used a value of $\sigma = 2.7$, which we found to be suitable for detecting most of the blobs in the image.

After the Gaussian filter is applied, we use the peak_local_max function to find the coordinates of the local maxima in the filtered image. This function works by first applying a maximum filter to the image, which finds the local maxima. The function then dilates the original input image, merging neighboring local maxima that are closer than the amount of dilation. The coordinates where the dilated and original images match are returned as local maxima.

With the coordinates of the local maxima established, we analyze the RGB levels of the central pixel of each blob to determine its color. We assign the color with the maximum value in the RGB codification to the blob. The software is designed to process frames in parallel, as there is no need to perform the analysis sequentially. If the number of detected coordinates does not match the expected number of kilobots, the corresponding image is placed in a separate folder for manual analysis. The source code for the software is available in GitHub [1]. Figure 9.2 shows an example frame of $N = 35$ kilobots *dancing* during one time step (A) and the same frame processed by the kilocounter software (B).

9.2 Experimental results

9.2.1 Distribution of bots detected over a time step

An interesting feature of opinion dynamics is the manner in which individual agents interact with each other in order to make the decision toward the available options. We first wanted to know how many kilobots an undecided bot detects during a time step Δt for our experimental conditions. As it will be further described in Section 9.2.2, the implemented model considers three different status. Dancing kilobots that move as PRWs, promoting either site 1 or site 2, and non-mobile undecided kilobots that are gathering information from their local environment to make a decision in the next time step. To prevent our group of kilobots from clustering at the wall of the circular observation area, the implemented PRW was configured with discrete wide turning angles so they would be able to turn away from the border faster. We established the turning times to either $\sim$ 2.8 seconds (turning angle of approximately

[1] https://github.com/TheFishTankLab/kiloColors

125°) or $\sim$ 5.8 seconds (turning angle of approximately 251°) and a forward motion time of $\sim$ 3.8 seconds.

Taking into account the limitations for kilobots to *express* their state, we first obtain information about the number of contacts of undecided kilobots with *dancing* kilobots by establishing ranges of bots detected, and mapping these numbers to a LED color. Thus, we implemented an algorithm for the undecided bots to communicate at each time step the range of bots they have seen in the previous time step. First, we documented the maximum number of kilobots detected by an undecided bot after several repetitions to calibrate the ranges of bots detected. We found that the maximum number was approximately 15 bots. We then divided the number of bots detected in four intervals, each one identified by an assigned color: red ($0-3$ bots), green ($4-7$ bots), blue ($8-11$ bots) and white ($12-15$ bots). At each time step, we counted the number of bots per color. We fixed an interaction radius restricted to $\sim$ 7 cm to ensure that bots would only detect a vicinity of no more than 2 kilobots' body lengths. We recorded five repetitions of 210 time steps (about 150 minutes), adding up to $19,195$ points. We extracted an image frame per cycle and counted colors with the *kilocounter* software described in Section 9.1.

Let $D_{n,m}$ be the ratio of kilobots that detected from n to m kilobots in a given time step Δt. Figure 9.3 shows a boxplot of $D_{n,m}$. Most bots received messages from 0 to 3 and from 4 to 7 bots in one cycle of $\sim$ 8.24 seconds, with mean values $\bar{D}_{0,3} = 0.710$ and $\bar{D}_{4,7} = 0.242$. The dispersion of the boxplots suggest that there is an heterogeneous spatial distribution across Δt cycles. Although in most cases bots are spread out and only come into contact with at most three neighbors, in some time steps, transient clustering of bots may enhance information exchange with a higher number of neighbor bots. Nonetheless, the small mean values for the higher ranges considered, $\bar{D}_{8,12} = 0.032$ and $\bar{D}_{8,12} = 0.016$ suggest that big clusters rarely happen, or that they dissolve rather quickly. From the behavior of $D_{n,m}$, we estimate the mean value of bots detected by an undecided bot to be of the order of 2.92 ± 2.50 during each time step Δt.

To extend our analysis of bots detected we also resorted to kilombo, the kilobots' simulator described in Chapter 7. First, we wanted to confirm that the number

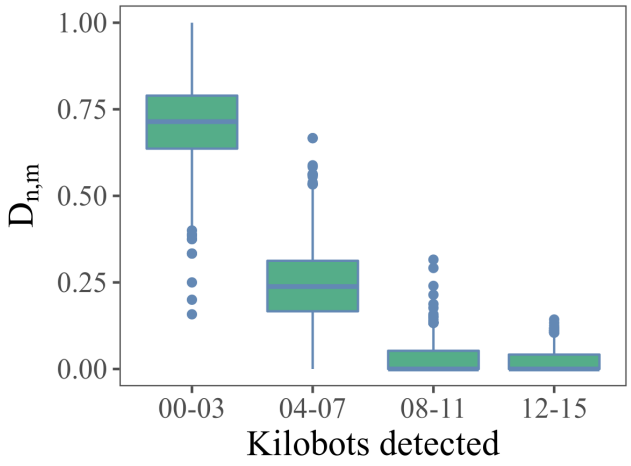


Figure 9.3: Bots detected in Δt in physical kilobots. The boxplots show the ratio of kilobots that detected from n to m kilobots in a given Δt cycle $D_{n,m}$ for four n, m ranges. Model parameters used for the exploration were $\lambda = 0$ and $\pi_i = \pi_j = 0.30$

of bots detected per time step Δt was similar to the one measured in physical kilobots. Second, kilombo makes it possible to extract more information that the bot has gathered during a decision-making cycle. Hence, we can obtain not only the total number of bots seen by an undecided bot during a time step Δt, but also the detailed fractions of bots detected that are dancing for each available site.

Figure 9.4 A shows the probability distribution of bots detected by kilobot i in a time step Δt, $B_{i,\Delta t}$, for the different values of the interdependence parameter λ considered in the experiments, both for the symmetric and asymmetric *a priori* discovery scenarios discussed in Chapter 8. Uncommitted bots in kilombo detected a mean of 2.91 ± 1.35 bots. The mean is the same as the estimated mean in physical kilobots, although in kilombo the distribution dispersion is more narrow. This discrepancy could represent the greater individual behavioral differences existing in physical kilobots that make up for the inherent noise of real life systems.

From our mean-field exploratory work in Chapter 8 we know that the population fraction promoting each site, f_j, varies greatly depending on the chosen model parameters. With this in mind, we analyzed the ratio of bots seen for each site by one individual during a time step Δt, $f_{i,j}$ for interdependence parameter values $\lambda = 0, 0.3, 0.6, 0.9$ in the case of symmetric *a priori* discovery probability values, $\pi_1 = \pi_2 = 0.3$, and asymmetric discovery probability values $\pi_1 = 0.40, \pi_2 = 0.20$, so that in the latter case, the site with low quality has a higher probability to be

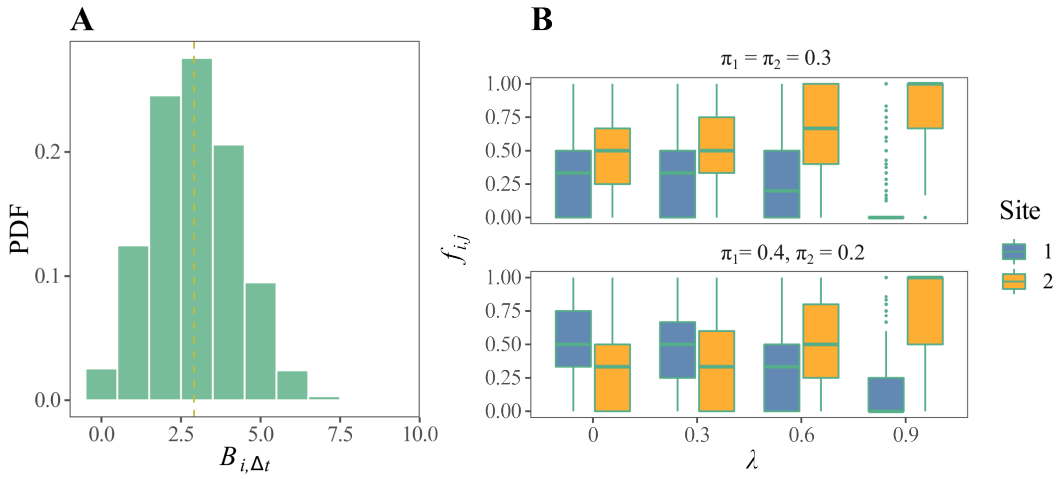


Figure 9.4: Bots detected during a time step Δt in kilombo. **A**: Probability distribution, $B_{i,\Delta t}$, of bots detected by a focal kilobot i in a time step Δt. **B**: Boxplots for the ratio of type j bots, bots dancing for site j, detected by a focal kilobot i in a time step Δt, $f_{i,j}$, as a function of the interdependence parameter λ for symmetric (up) and asymmetric (bottom) *a priori* discovery probabilities. Data aggregated over 50 repetitions per parameter combination.

found. We fixed site qualities to $q_1 = 7$ and $q_2 = 10$, which translate to dance times of $7\Delta t$ for the low quality site, and $10\Delta t$ for the high quality site. These values were chosen so that we could observe a transition from non-consensus to consensus as a function of λ. Remember from Chapter 8 that for big quality differences, i.e. $q_1 < 6$ for $q_2 = 10$, the robot swarm achieves consensus for all λ's, while close site quality values hinder consensus at lower values of λ.

Let $f_{i,j}$ be the ratio of type j bots, i.e. bots dancing for site j, over the number of bots detected in a time step by a focal individual i. Figure 9.4B shows boxplots for $f_{i,j}$ as as a function of the interdependence parameter for both symmetric and asymmetric discovery probabilities π_j's. In the boxplots, we can observe the spread and skewness of the distribution $f_{i,j}$ for each available site. The dispersion of $f_{i,j}$ is large for both sites and discovery probabilities combinations, specially for intermediate λ's. We can also see a clear increasing trend for non-dancing bots detecting higher numbers of bots dancing for site 2, the high quality site, as λ increases, whereas bots dancing for site 1 are seldom encountered as λ increases. Moreover, as expected, for the asymmetric discovery probabilities case, the difference between seen-bots distributions for the two sites available is narrower than for the symmetric case at each λ, as the site of less quality has a higher rate of independent discovery.

9.2.2 Consensus reaching in a group of kilobots

After exploring the distribution of contacts among kilobots during a time step Δt, we now proceed to analyze the List et al., 2009 decision making process in our kilobots experiment.

At each time step, kilobots are either dancing for site 1 (the low quality site), site 2 (the high quality site) or not dancing for any site. Kilobot's LED lights changed color to show each bot state: red for undecided bots, green for bots dancing for site 1 and blue for bots dancing for site 2, as shown in Figure 9.1 After a new time step has passed, kilobots that were dancing for a site either continue their dance if there is dance time left, or they stop if the dance time is over. Those that were not dancing updated their probability to start promoting a site based on Eq.8.1, with f_i the frequencies calculated by every non-dancing bot over the time interval Δt, i.e. the ratios $f_{i,j}$ discussed in the previous section. Kilobots that are not dancing for any site do not move, while other bots promote the site they dance for by performing a PRW motion such as the one described in Chapter 7.

Here we analyze the experimental behavior of dance frequencies and of consensus values for different levels of interdependence for two particular cases with symmetric and asymmetric discovery probabilities. In particular, we recorded five repetitions of the experiment for interdependence parameter values $\lambda = 0, 0.3, 0.6, 0.9$, both for the symmetric discovery probability values $\pi_1 = \pi_2 = 0.30$, and for the asymmetric discovery probability values $\pi_1 = 0.40$, $\pi_2 = 0.20$. In total, these recordings represent approximately 20 hours of tracking the kilobots decision making process. From the recordings, we extracted an image frame per cycle and counted colors with the kilocounter software [2]. From the extracted data, we analyzed the temporal evolution and obtained stationary values for dance frequencies f_j, and for the strong consensus $Q = f_2 - 2f_1$, defined in Chapter 8, for each parameter combination.

Temporal evolution of dance frequencies

Figure 9.5A shows dance frequencies f_0, f_1, f_2 as a function of time, averaged over five repetitions. Our initial conditions are $f_1(t = 0) = f_2(t = 0) = 0.43$ for bots

[2] https://github.com/TheFishTankLab/kiloColors

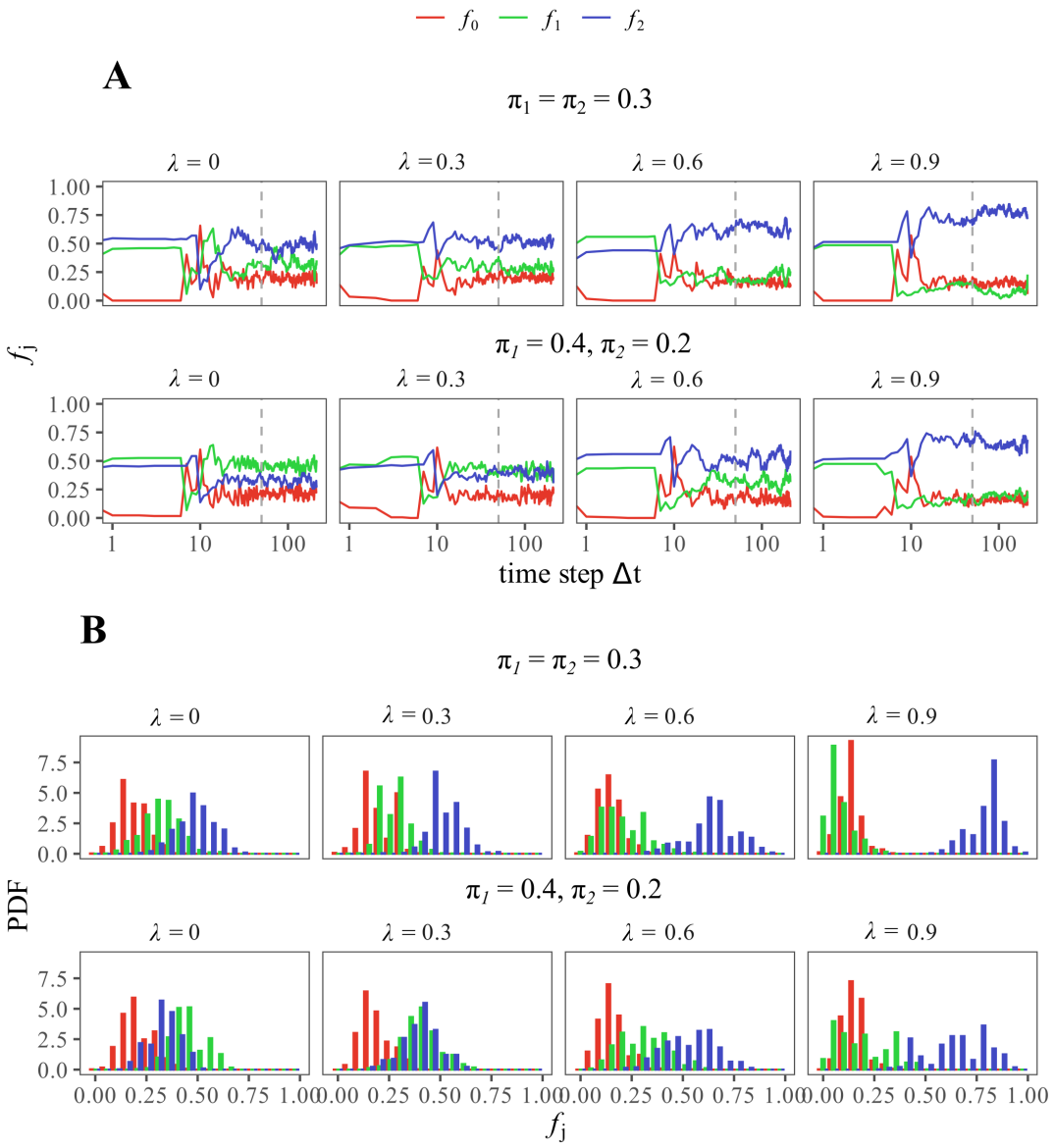


Figure 9.5: **A**: Dance frequencies f_j as a function of time (averaged over five repetitions). In red, green, and blue, we represent the fraction of uncommitted, type 1, and type 2 bots, respectively. **B**: Histograms of f_j values in the stationary state. Plots correspond to symmetric (up) an asymmetric (bottom) *a priori* discovery probabilities, and to interdependence parameter values $\lambda = 0, 0.3, 0.6, 0.9$ (left to right).

dancing for either site 1 or site 2, and a lower probability of $f_0(t = 0) = 0.14$ of bots that start undecided. It is noticeable that, as it is also observed in mean-field simulations (see Figure 8.1 in Chapter 8), an increasing interdependence broadens the difference between the frequency of bots dancing for low and high quality sites. Increasing the value of the interdependence parameter increments f_2 and reduces f_1 and, to a lesser extent, f_0. This becomes clearer in kilobots' dance frequency distributions during the stationary state, achieved after approximately 50 time steps, as

shown in Figure 9.5 B. All conditions have a large f_j dispersion. For low values of λ all three distributions overlap, but as λ increases f_1 and f_2 distributions separate until they have different mean values at the highest value of $\lambda = 0.9$. Nonetheless, few values can still overlap at some time steps due to its recurrent state updating, a characteristic feature of the model considered.

Indeed, an inherent consequence of the model dynamics is that bots promoting a site return to an uncommitted state when their dancing time is over, thus continuously changing the decided-undecided ratio. This property causes the dance frequencies obtained for kilobots to fluctuate in time. Finite size effects amplify these fluctuations even further. Thus, the stationary state is characterized by a peaked-distribution around a central mean value and an associated standard deviation, as discussed in the model parametric exploration presented before in Chapter 8. These relatively large fluctuations also make it difficult to quantitatively determine a characteristic transition time towards the stationary state. For practical purposes, we set a minimum number of time steps after which we can assume that our system evolution is already stationary. In these experiments, we fixed this transient time to be of the order of $50\Delta t$, which is shown as dashed gray lines in Figures 9.5 A. Although stationary values do not depend on the initial preparation of the system, this transient time, obviously, does.

Stationary dance frequencies and consensus reaching in kilobots

After getting a sense of the temporal evolution of dance frequencies $f_i(t)$ in kilobots, in this section we analyze the average behavior of stationary dance frequencies $\langle f_j \rangle$ and the corresponding stationary consensus values $\langle Q \rangle$ as a function of the interdependence parameter in the model λ. Averages are obtained after five repetitions per experimental condition in our kilobot ensemble. Figures 9.6 summarize our main experimental results. For completeness, these results are also compared with simulation results as explained in detail in the following paragraphs.

Firstly, the same model and conditions are also implemented in the kilombo simulator for proper comparison with the experimental results. In kilombo simulations, we have considered a speed value of $9 \pm 2\ mm/s$ to match the movement of physical kilobots. As explained before, kilombo simulations allow us to scale up the system

and to improve our statistics far beyond our current experimental capabilities. Here, we represent the average behavior $\langle f_j \rangle \pm S.D.$ and $\langle Q \rangle \pm S.D.$ averaged over 50 realizations of the kilombo simulations for each experimental condition. Secondly, we also include the mean stationary values of the same quantities averaged over 100 realizations of the mean-field approximation, described in the previous Chapter 8, again for the sake of providing an experimental-theoretical comparison.

Finally, we also included the comparison with stationary values of dancing frequencies and consensus parameter for a *quenched* approximation of the same collective decision making model. The latter replaces real or simulated kilobots by immobile agents randomly deposited inside the circular arena of radius R, as described in Chapter 7 and shown in Figure 7.11 B. Agents are kept quenched at these same positions through the whole evolution of the model. Moreover, dancing probabilities p_j are calculated after exchanging information with neighboring particles located within an interaction radius r_i. As in the experimental set up, here we use $N = 35$ bots, an arena of radius $R = 20$ cm , as well as an interaction radius $r_i = 5$ cm and an exclusion radius $r_e = 1.5$ to match our kilobots' interaction characteristics and body size, respectively.

The quenched approximation is included to allow us to compare our experimental results to this immobile limit, where the mixing of agents is completely absent and the locality of kilobots interactions should prevail, and thus depart as much as possible from the mean-field, all-to-all interaction, approximation.

π_1	π_2	λ	$\langle f_2 \rangle$	$\langle Q \rangle$
0.3	0.3			
		0	0.478 ± 0.088	-0.171 ± 0.25
		0.3	0.524 ± 0.079	-0.062 ± 0.2
		0.6	0.634 ± 0.113	0.219 ± 0.319
		0.9	0.775 ± 0.077	0.63 ± 0.197
0.4	0.2			
		0	0.325 ± 0.078	-0.586 ± 0.221
		0.3	0.398 ± 0.09	-0.419 ± 0.251
		0.6	0.522 ± 0.123	-0.101 ± 0.341
		0.9	0.644 ± 0.144	0.26 ± 0.401

Table 9.1: Stationary $\langle f_2 \rangle$ and $\langle Q \rangle$ values for physical kilobots averaged over five repetitions.

Figures 9.6 A and B show the experimental average dancing frequencies $\langle f_j \rangle$ and

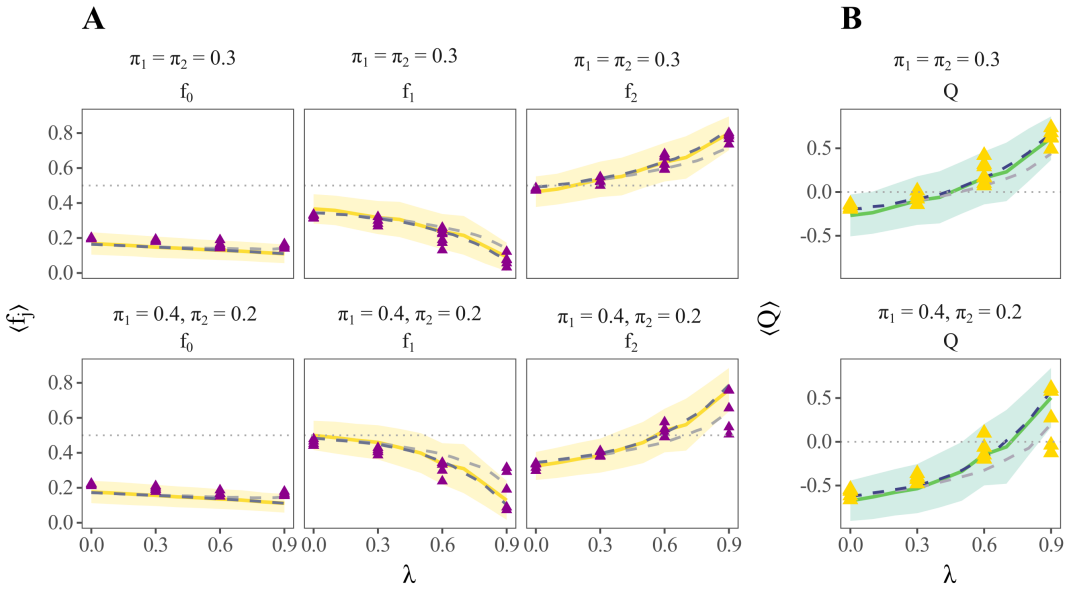


Figure 9.6: **A**: Mean Dance frequencies $\langle f_j \rangle$ as a function of λ for symmetric (up) and asymmetric (bottom) π_j for kilobots, kilombo, mean field and quenched simulations. **B**: Mean consensus parameter $\langle Q \rangle$ as a function of λ for symmetric (up) and asymmetric (bottom) *a priori* discovery probabilties π_j for physical kilobots, kilombo, mean-field and quenched simulations. **Triangles**: Experimental kilobots (five repetitions per parameter combination). **Solid line**: Mean values obtained from kilombo simulations (averaged over 50 repetitions). Shaded area: Standard deviation in kilombo simulations. **Dashed dark line**:Mean values obtained in mean field simulations (averaged over 100 realizations). **Dashed light line**: Mean values obtained from numerical simulations in the quenched approximation (averaged over 100 realizations).

consensus parameter $\langle Q \rangle$ together with the three different approximations we have just described. From the Figures we can generally observe that the resulting stationary values in physical kilobots are very similar, within the shaded error area, to kilombo simulations. Moreover, experimental results also fall close to mean field and to quenched approximation results. The latter finding is not so surprising when one realizes that, for the parameters characterizing our experimental setup and the swarm of physical kilobots, that both approximations (the mean-field and the quenched approximations) are also close to each other, as can be observed in these figures. Below, we will further discuss and justify this, *a priori*, unexpected result, but first let's examine the kilobots' stationary behavior.

For the parameters considered in the experimental tests, the average value $\langle f_1 \rangle$ decreases whereas the average value $\langle f_2 \rangle$ increases as interdependence becomes more relevant. Additionally, both dance frequencies exhibit a similar functional trend but in opposite directions, with average stationary $\langle f_2(\lambda = 0.9) \rangle - \langle f_2(\lambda = 0) \rangle = 0.30$ and average stationary $\langle f_1(\lambda = 0.9) \rangle - \langle f_1(\lambda = 0) \rangle = -0.25$ for the case

of symmetric discovery probabilities, and $\langle f_2(\lambda = 0.9)\rangle - \langle f_2(\lambda = 0)\rangle = 0.32$ and average stationary $\langle f_1(\lambda = 0.9)\rangle - \langle f_1(\lambda = 0)\rangle = -0.26$ for asymmetric discovery conditions. Thus, the stationary value of $\langle f_0\rangle$ remains almost constant. For example, it only decreases by a small amount $\langle f_0(\lambda = 0.9)\rangle - \langle f_0(\lambda = 0)\rangle = -0.045$ for the symmetric discovery case for the physical kilobots. The continuous introduction of independent discoveries for both sites reduces the amount of undecided bots. If $\pi_j \to 0$, f_0 would be higher, specially for $\lambda \to 0$ and for short dancing times, i.e. $q_j \to 0$.

Regarding the stationary average of our strong consensus parameter $\langle Q\rangle$ introduced in Chapter 8, in Figure 9.6B, we can observe that there is a smooth crossover from non-consensus to reaching a consensus as a function of the interdependence parameter λ. Moreover, consensus is higher for the symmetric discovery probabilities scenario, as there is less introduction of independent discovery of the low quality site, than in the asymmetric case.

Kilombo simulations reproduce the physical kilobots' behavior: most stationary values of the kilobots' fall within the standard deviation shaded area of kilombo. These results indicate that improving the statistics in the physical kilobots experiments would eventually yield better agreement with average kilombo results.

Remarkably, experimental results from physical kilobots closely match stationary values predicted by the mean-field approximation. In the case of kilombo simulations, stationary values are found to overlap with mean-field stationary values under all analyzed conditions. These findings suggest that mobile agents mixing with their local environment can effectively sample the state of the system and transmit information throughout, allowing for the attainment of collective decision-making comparable to that of mean-field connected agents that have full access to group information for individual decision-making.

To explain the observed similarity, it is necessary to delve into our quenched approximation. As previously mentioned, we also find that quenched bots configurations, featuring the same characteristics as physical kilobots, yield very similar stationary dance frequencies than physical kilobots and mean-field results. Indeed, all these results are almost identical for low values of λ, and they start differing after an interdependence parameter value of approximately $\lambda > 0.5$, when $\langle f_1\rangle$ is

higher for quenched simulations and $\langle f_2 \rangle$ attains lower values. This, in turn, results in a lower consensus value (Figure 9.6B), and even in a no-consensus situation in a region of λ values where kilobots and mean-field simulations could achieve it, especially for the asymmetric discovery case.

Thus, the fact that communication occurs between all agents (i.e. mean-field interactions) only implies a better consensus parameter in the region where the interdependence parameter λ is higher than a given threshold, which is higher in the symmetric discovery scenario than in the asymmetric case. Quenched configurations with fixed neighbors interactions emphasize the relevance of kilobots mixing under the experimental conditions considered in our setup. Despite the limited range of kilobots' interactions, their mobility enables information to travel in a way that is similar to mean-field results. Quenched conditions, on the other hand, can promote low-quality environments where low-quality sites can dominate the consensus dynamics.

Low interdependence reduces the impact of agents' communication in the individual decision making process. As interdependence increases, communication becomes more relevant for consensus reaching. Within the high interdependence regime, our results show that the mixing agents, i.e. kilobots moving in space through time, is indeed needed for reaching higher consensus values as the ones obtained in mean field simulations.

One possible explanation of the fact that quenched stationary values are lower for high interdependence than in kilombo is that the distribution of stationary states in quenched approximations is more dispersed. This could be attributed to the heterogeneity of the agents' spatial distribution between different simulation realizations in the quenched approximation. With that in mind, we next analyze the dispersion of our stationary data for all of our approximations. In Figure 9.7 we show the standard deviation of stationary dance frequencies and consensus.

The impact of interdependence on the values of the stationary states is noticeable. When the interdependence parameter λ is set to 0, the dispersion of stationary values could be considered as the inherent variation in the system when decisions are made independently (see the analytical description of the model provided in the

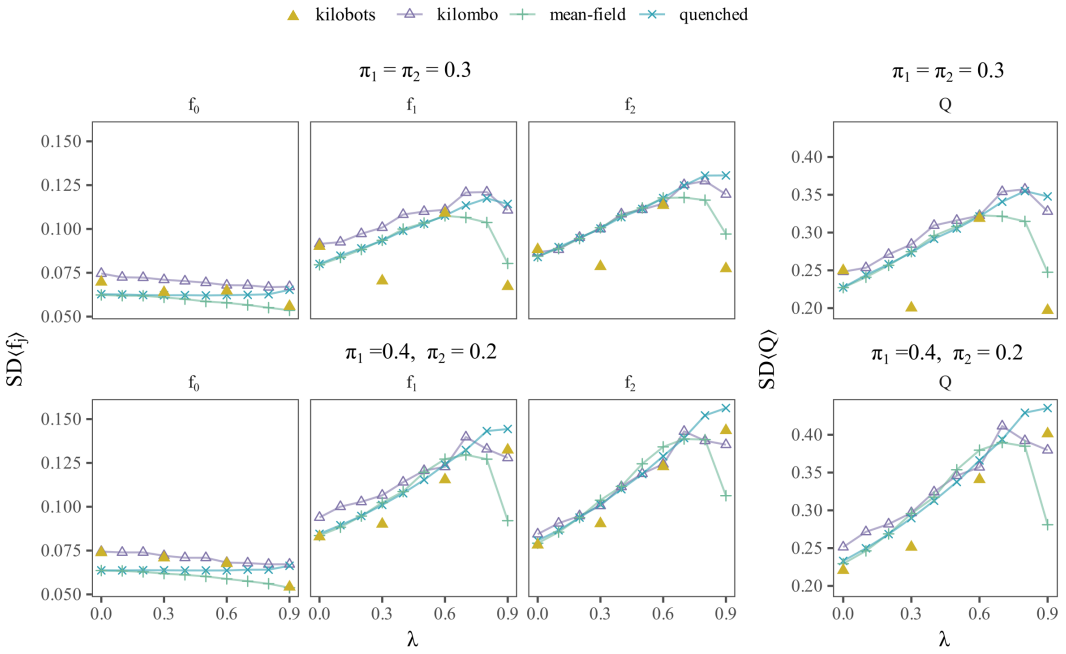


Figure 9.7: **A**: Standard deviation of $\langle f_j \rangle$ as a function of λ for symmetric (up) and asymmetric (bottom) discovery scenarios for physical kilobots, kilombo, mean field and quenched simulations. **B**: Standard deviation of $\langle Q \rangle$ as a function of λ for symmetric (up) and asymmetric (bottom) discovery probabilities Π_j.

previous Chapter 8). The continuous change in the number of dancing and non-dancing agents, as well as the probabilistic nature of the model, contribute to this dispersion.

As interdependence increases, the standard deviation of f_1, f_2 and Q grow in a similar way in all three types of simulations (mean field, quenched, and kilombo) for a low to medium λ range. However, for $\lambda > 0.6$, there is a clear difference between mean field and quenched simulations. In this range, the mean field standard deviation decreases significantly, while the quenched standard deviation continues to increase. That is, the fact that agents calculate the probabilities of dancing for a site taking into account all other agents' opinions in the mean-field approximation increases the *certainty* in the consensus value and the consensus itself is stronger.

The higher dispersion in quenched approximations may result from the variations in the static spatial distributions of agents across repetitions of the simulations, which affects local interactions among neighboring bots. In situations where agents rely heavily on the opinions of their neighbors, it can lead to local environments

where agents are promoting different sites. This can amplify contrasting dance frequencies around the group and cause the value of consensus to vary among repetitions.

This is especially pronounced as independent discovery probabilities are higher and low quality sites are easier to find, as we observe for $\lambda = 0.9$ in our asymmetric case. Kilombo simulations show a sort of intermediate standard deviation value between mean field and quenched approximations for $\lambda > 0.7$. The number of repetitions performed in the physical kilobots probably does not represent the entire stationary values distribution. As such, kilobots' standard deviation behavior when increasing interdependence is noisier than kilombo, but in most cases it follows the same trend, particularly for asymmetric *a priori* probabilities. Kilobot's local interactions and mixing through time do achieve as high a consensus as mean field approximations, but the probability of obtaining a different value of consensus is higher for kilobots.

9.3 Cluster sizes and percolation of information spreading

Having analyzed the variations in dance frequencies and consensus across our multiple approximations, in this next section, we aim to investigate why collective decision-making in physical kilobots is as effective as in our idealized mean field approximation. One possible explanation is that, despite detecting only about 2.9 other bots, on average, the movement of kilobots and the system's overall density enable the formation of a connected network of bots, which facilitates the transmission of information throughout the system. In this communication network, bots receive information beyond their immediate surroundings, even from other bots that are not directly detected by their sensors, through the percolation of their intercommunication areas in continuous space.

The kilobot contact network (Newman, 2010) can be constructed in terms of a geometrical network (Barthélemy, 2011), in which nodes represent kilobots at a particular position in space and two nodes are connected by an edge if the corresponding kilobots are at a Euclidean distance smaller than their interaction radius. As communication in kilobots is approximately isotropic, the network is undirected, i.e. we

assume that if bot i interacts with bot j, bot j would also interact with bot i. This network has a time-varying nature, since the configuration of contacts changes as the kilobots change position in the experimental arena.

Here, we aim to investigate these issues by extending our analysis to examine the characteristics of the kilobots' contact network. Understanding physical quantities such as the cluster size distributions, the average connectivity, etc. for various system sizes, interaction radii, and exploratory time windows will prove crucial in understanding our results.

In the following subsections, we characterize the connectivity properties of a kilobot ensemble, both in the quenched approximation and under dynamics conditions such as the one implemented in the experiments, through the kilombo simulator. The use of the simulator is fundamental in this respect since it allows us to scale up the size of our system to understand the fundamental ongoing physical processes, and therefore, to reduce the unavoidable spurious finite size effects inherent to our small group of physical kilobots.

9.3.1 Mean cluster size

We consider the cluster structure of the kilobot contact networks constructed from the vast amount of spatial configurations generated in kilombo simulations. A cluster in the kilobot network is defined as a connected component in which all pairs of nodes can be reached from one to the other following a continuous path of adjacent edges (Newman, 2010). From a computational perspective, we will say that a bot belongs to the same communication cluster of a focal bot if it is located within a circular region of radius r_i, the effective interaction radius, centered on the focal bot. A third kilobot will belong to the same cluster when its separation from any one of the previous bots is shorter than the interaction radius r_i. The process is iterated over all bots to finally determine all the clusters, as well as their size (the number of components), present for each spatial configuration of the system. A cluster in this context can thus be understood as a group of kilobots that are located within their infrared inter-communication distance, or interaction distance, estimated from pairs of neighboring kilobots.

The basic characterization of the cluster structure is given by the mean cluster size $\langle \mathcal{S} \rangle$, which is computed as (Stauffer and Aharony, 2018)

$$\langle \mathcal{S} \rangle = \frac{\sum' s^2 n(s)}{\sum' s n(s)},$$

where $n(s)$ is the number of clusters of size s (composed by s bots) and the summation in the denominator does not include the giant component of the network, i.e. the largest cluster observed in a given bot configuration. The result is averaged over all configurations comprising our statistics.

The mean cluster size is a fundamental quantity in the geometrical process of percolation (Stauffer and Aharony, 2018), as it plays the role of a transport coefficient, such as the magnetic susceptibility or the specific heat, quantifying the fluctuations of the cluster size distribution. In the context of our contact networks, it can be used to detect the presence and position of a percolation transition, in which the communication network passes from having only small isolated clusters of bots to having a large connected communicating component. At fixed density, this transition takes place as a function of the interaction radius r_i, and is characterized by a threshold value at which the mean cluster sizes diverges in the thermodynamic limit. In the case of our finite experimental networks, the mean cluster size does not show a divergence, but a maximum at the position of the interaction radius threshold (Stauffer and Aharony, 2018; Newman, 2010). Studying the evolution of the position and height of this maximum, applying a finite-size scaling analysis, allows to obtain numerical information about the transition properties (Stauffer and Aharony, 2018).

In Figure 9.8A, we plot $\langle S \rangle$ for different system sizes, preserving the same number density $N/\pi R^2 = 0.028$ bots$/cm^2$, in the quenched approximation. This figure clearly illustrates the effects of system size for different values of the interaction radii r_i, which characterizes the maximum extent of message transmission, and thus of information exchange, through infrared sensors among physical kilobots. From kilombo data in the quenched approximation, we can identify the critical percolation interaction radius at around $r_i^* = 6.5 \pm 0.2$ cm for $N = 35$ and $r_i^* = 6.4 \pm 0.2$ cm for $N = 492$. Continuous percolation threshold values for two dimensional discs of effective radius r_i in a square box of linear dimension L and periodic boundary

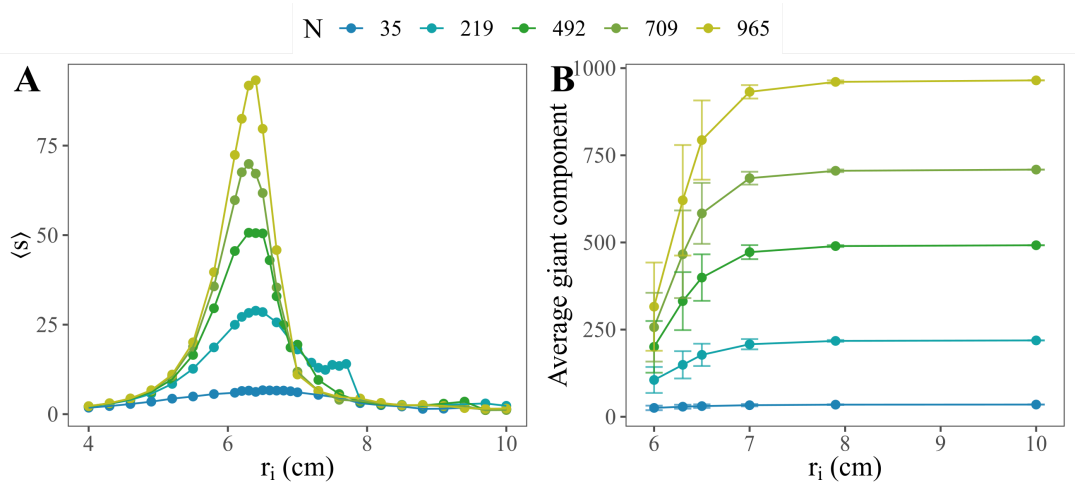


Figure 9.8: **A**: Mean cluster size as a function of the interaction radius r_i for quenched configurations with different system sizes, preserving the same number density. **B**: Average giant component as a function of r_i for quenched configurations with different system sizes, preserving the same number density. Cluster sizes were averaged over 1000 configurations per system size.

conditions are found in the literature (Mertens and Moore, 2012). The critical filling factor in that particular geometry is $\eta^* = \pi r_i^2 N/L^2 \simeq 1.128$, or in other words, $r_i^* = (L^2\eta^*/(N\pi))^{1/2}$. Assuming a similar scaling behavior in our case, with a fixed rigid circular wall, would yield a smaller threshold radius of approximately 3.59 cm, indicating that both our circular geometry and fixed boundary conditions give rise to packing and size effects that cannot be neglected in the quantitative determination of this non-universal threshold value. On the other hand, such effects should not be relevant regarding the behavior of critical exponents. In Figure 9.8B, we also represent the average size of the so-called *giant component*, or the size of the largest connected cluster in the system, as a function of the interaction radius r_i. This quantity attains its maximum value, comparable to the system size, after the percolation threshold. As the maximum value of the mean cluster size, the size of the *giant component* at the percolation threshold scales as a power law of the system size. Likewise, at the percolation threshold, the probability distribution of cluster sizes $P(s) = n(s)/n_{tot}$ represented in Figure 9.10, decays algebraically as a power law up to a given limiting size determined by finite size effects, where this probability density decays much faster towards zero. For finite system sizes, or near the critical point, only finite clusters exist up to a largest cluster size, and the cluster-size distribution is smoothly cut off by a rapidly decaying function. A proper finite-size scaling analysis of all these

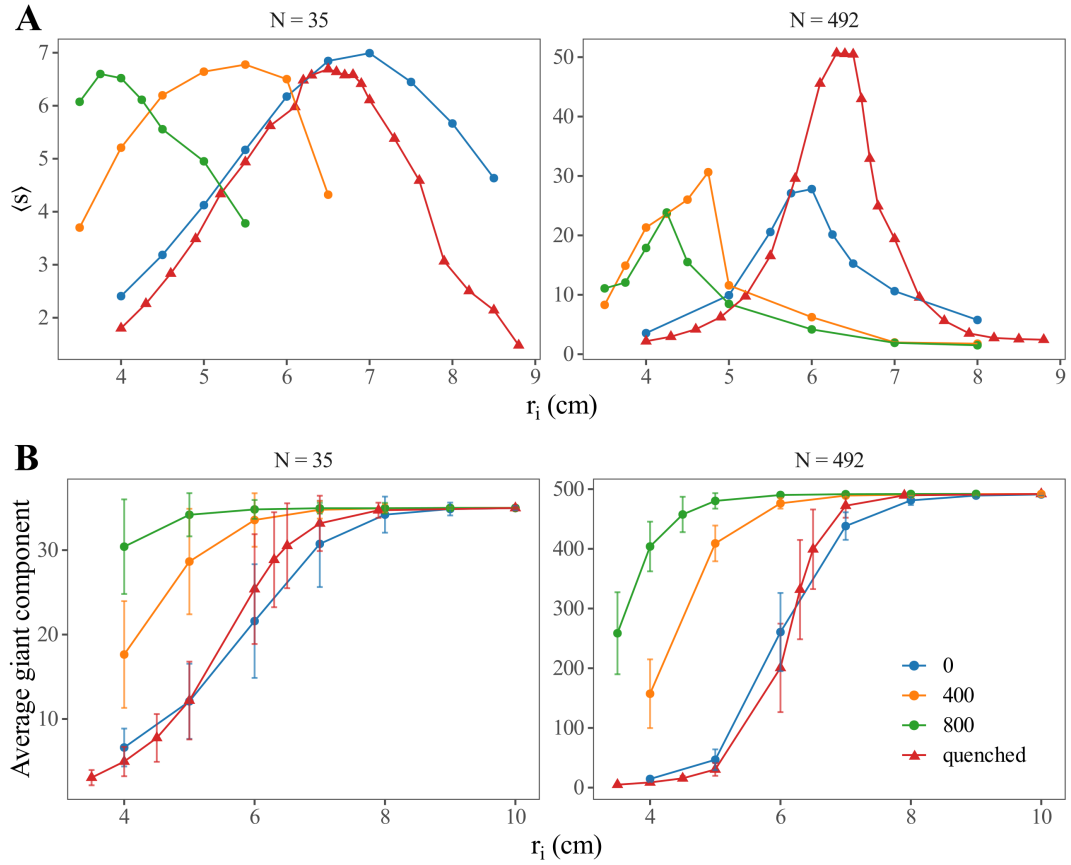


Figure 9.9: **A**: Mean cluster size as a function of the interaction radius r_i for quenched configurations and kilombo simulations at different integration times $\Delta t = 0, 400, 800$ kilobot iterations. **B**: Average giant component as a function of r_i for quenched configurations and kilombo simulations at different integration times $\Delta t = 0, 400, 800$ kilobot iterations. The left panel shows the results for system size $N = 35$, while the right panel shows the results for system size $N = 492$.

power-law behaviors will provide critical exponent values in agreement with previous results published in the literature (Stauffer and Aharony, 2018). However, such detailed finite-size scaling analysis is, however, beyond the scope of this thesis.

In addition, if we now consider the PRW trajectories performed by kilobots in the two dimensional kilombo simulations, we can also compute the mean cluster size of their communication network integrated over their exploratory time window Δt, i.e. of the total number of transmission contacts accumulated during their exploratory time step Δt. In Figure 9.9 we now represent the mean cluster size obtained for different exploratory time windows of $\Delta t = 0, 400, 800$ kilobot loop iterations or, equivalently, of $\Delta t = 0, 4.12, 8.24$ seconds, and different system sizes with the same number density of bots $n = N/\pi R^2 = 0.028\,\text{bots}/cm^2$.

A temporal window of $\Delta t = 0$ should match the results obtained from the

quenched configuration approximation described in the previous paragraphs, whereas increasing values of Δt would yield increasingly larger communication cluster sizes as a result of the kilobot exploratory dynamics. Indeed by moving, bots increase their average number of communication contacts favouring the widespread of information through the system. Quantitatively speaking, this will eventually translate into the reduction of the percolation threshold radius above which we can find a giant communication component that allows the exchange of information among almost all bots in the system. In Figures 9.9 and 9.10, we represent our main results for the mean cluster size, the *giant component* size, and the distribution of cluster sizes, in comparison with the results of the quenched approximation.

As expected, for large enough system sizes, the position of the peak of the mean cluster size for $\Delta t = 0$ is compatible, within error bars, with that corresponding to the quenched approximation. This implies that potential spatial correlations originating from the accumulation of some bots against the arena wall, or from bot collisions, are unimportant for large enough system sizes. These correlations are, of course, absent in the quenched positions approximation. Although the qualitative behavior remains similar for small system sizes, spatial correlations induced by the motion of kilobots near the wall of the arena may generate spurious size fluctuations for interaction radii larger than the percolation threshold, spurious fluctuations that disappear for larger system sizes.

In Figures 9.9A and B, one can also observe the shift of the mean cluster size distribution peak, as well as the average size of the *giant-component*, towards smaller values of r_i^* as we increase the exploratory time window Δt. For the larger system size of $N = 492$ bots, the percolation threshold moves from the quenched approximation value around $r_i^* = 6.4 \pm 0.2$ cm to the smaller value of $r_i^* = 4.75 \pm 0.25$ cm for $\Delta t = 4.12$ s (400 loops), and to an even smaller value around 4.25 ± 0.25 cm for the exploratory time window of $\Delta t = 8.24$ s (800 loops). The same observations hold true for the small system size of $N = 35$ bots, for which the percolation threshold for the quenched approximation is around 6.5 ± 0.1 cm, and the threshold value after an exploration time window of $\Delta t = 8.24$ seconds (800 loops) is located at $r_i^* = 3.75 \pm 0.25$ cm, rather close to the kilobot excluded distance of $d_e = 3.0$ cm. In addition, at the corresponding finite size percolation thresholds, the distributions

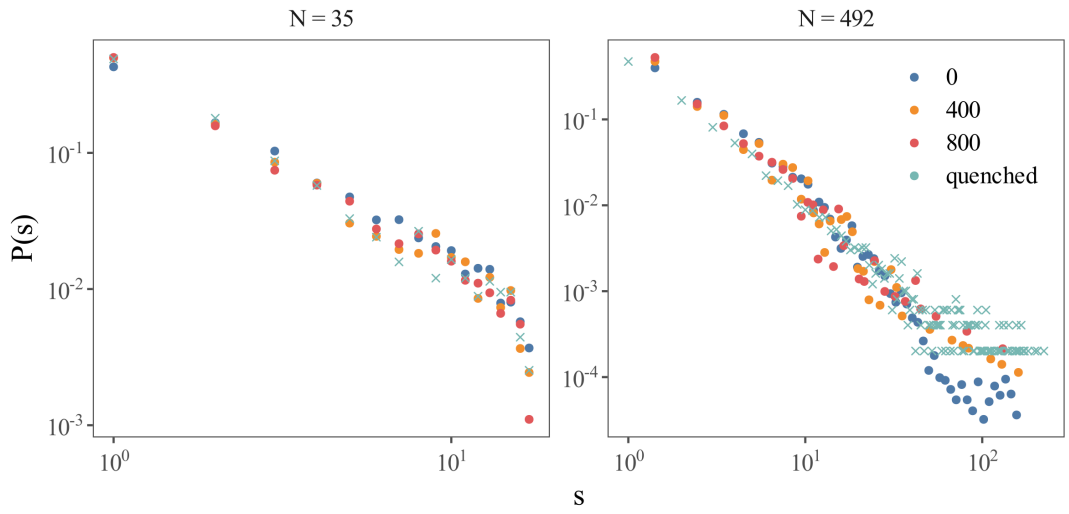


Figure 9.10: Probability distribution of cluster sizes $P(s)$ at the percolation transition for quenched configurations and kilombo simulations at different integration times. The left panel shows the results for system size $N = 35$, with percolation transitions at the following r_i^*: Quenched ($r_i^* = 6.4$ cm), $\Delta t = 0$ ($r_i^* = 7.0$ cm), $\Delta t = 400$ ($r_i^* = 5.5$ cm), and $\Delta t = 800$ ($r_i^* = 3.8$ cm). The right panel shows the results for system size $N = 492$, with percolation transitions at the following r_i^*: Quenched ($r_i^* = 6.5$ cm), $\Delta t = 0$ ($r_i^* = 6.0$ cm), $\Delta t = 400$ ($r_i^* = 4.8$ cm), and $\Delta t = 800$ ($r_i^* = 4.3$ cm)

of cluster sizes depicted in Figure 9.10, decay as a truncated power law of s with compatible exponent values within error bars.

Finally, it is worth emphasizing the latter results since the mobility of kilobots over the exploratory temporal window Δt can readily ensure the percolation of the communication network for relatively short communication radii r_i, radii slightly larger than the kilobot radius. Remember that physical kilobots in our kilombo experiments have an interaction radii of 5 ± 2 cm, and the possibility to explore their neighborhood for a time window of $\Delta t = 8.24$ seconds (800 loops) . Therefore, from this point of view, kilobots form a percolating communication network that allows them to exchange information at the system size level, or in other words, as in a mean-field interaction approximation.

9.3.2 Communication network degree and degree distribution

In network theory, the degree of a node k is defined as the number connections or edges that the node has to other nodes, while the degree distribution $P(k)$ is defined as the probability that a randomly chosen node has degree k (Newman, 2010). The

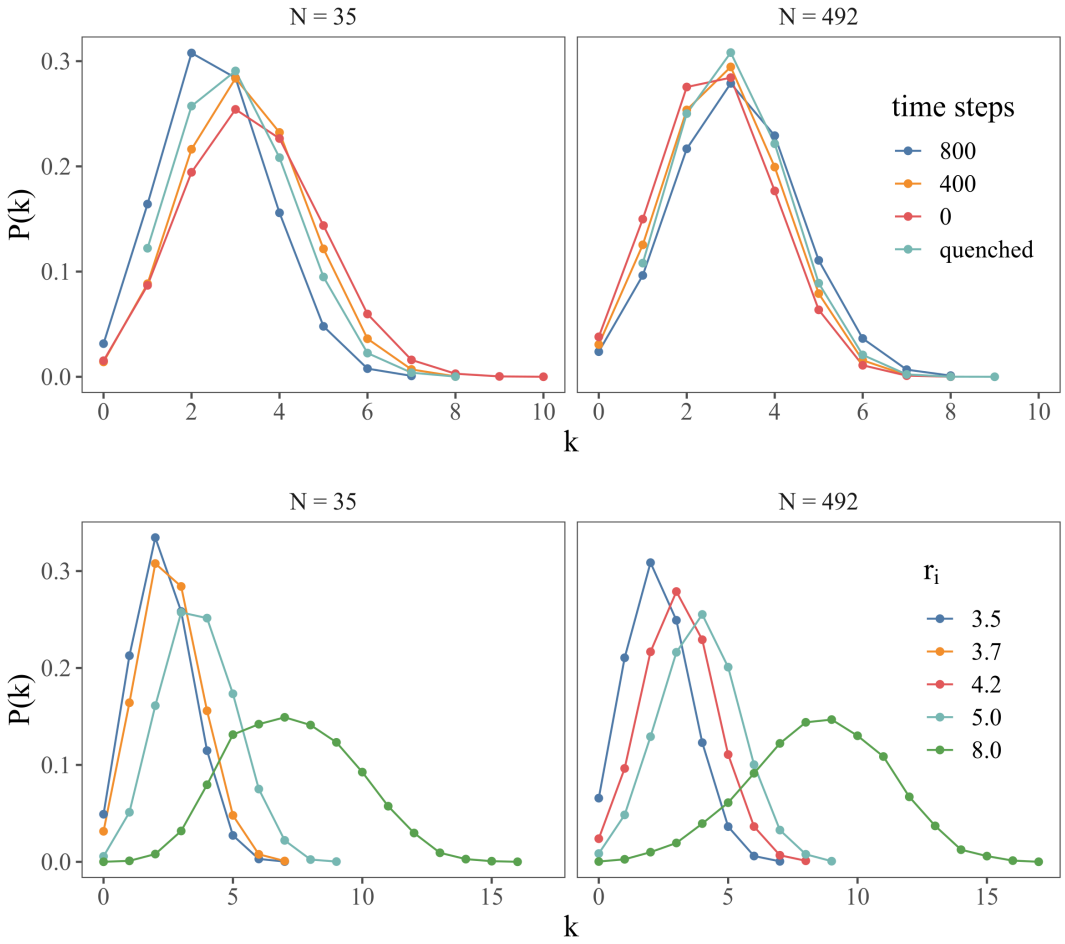


Figure 9.11: Degree distribution $P(k)$ in kilobots in kilombo for different time steps of $\Delta t = 0, 400, 800$ loop iterations at the threshold r_i^* and for 1000 quenched configurations. (top panels). Bottom panels represent results for different interaction radius r_i and $\Delta t =$ 800 loop iterations. Left panel correspond to $N = 35$ and the right panels to $N = 495$.

degree and the degree distribution are important factors in determining the properties of dynamical processes taking place on networks (Barrat et al., 2008), such as in this case, the spread of information through an infrared communication network.

In Figure 9.11 we show the degree distribution observed in the kilobot communication network obtained from kilombo simulations over different values of the exploratory time step Δt, and different radius of interaction r_i, for two different system sizes with the same number density of kilobots. From the figure we can see that the distributions are bell-shaped, roughly compatible with a Poisson form,

$$P(k) = e^{-\lambda} \frac{\lambda^k}{k!}, \tag{9.1}$$

with λ equal to the average degree $\langle k \rangle$.

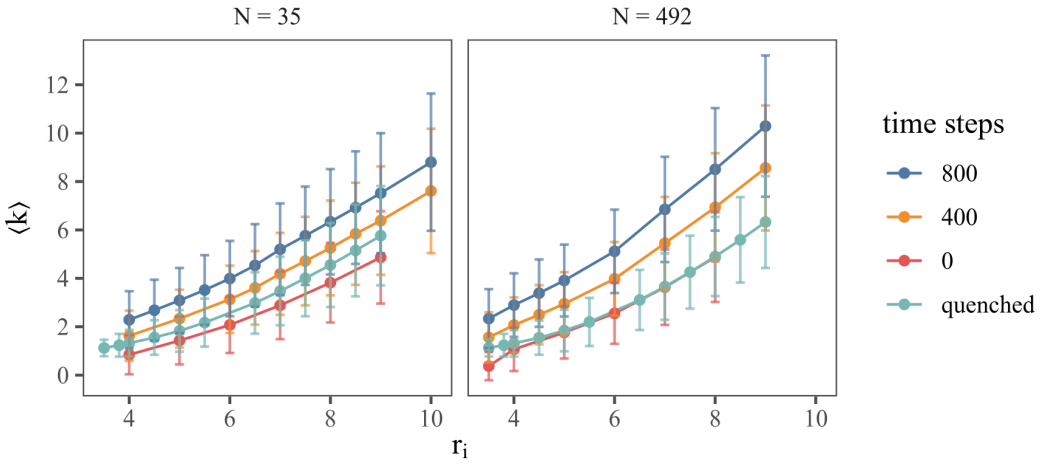


Figure 9.12: Integrated average degree $\langle k \rangle$ as a function of r_i in kilobots in kilombo for different time steps of $\Delta t = 0, 400, 800$ loop iterations, and for 1000 quenched configurations. Left panel shows $\langle k \rangle$ for $N = 35$ and right panel for $N = 495$.

As in the previous subsection, we have also calculated the integrated average degree of the kilobot communication network, denoted by $\langle k \rangle$, in kilombo simulations over different values of the exploratory time step Δt, and two different system sizes with the same number density of kilobots. Figures 9.12 show the integrated average degree as a function of r_i for the same exploratory time steps of $\Delta t = 0, 400, 800$ kilobot loop iterations or, equivalently, of $\Delta t = 0, 4.12, 8.24$ seconds, for $N = 35$ and $N = 492$ bots (and the same number density $n = N/\pi R^2 = 0.028$ bots$/cm^2$) in kilombo simulations. For comparison, we also include $\langle k \rangle$ values for 1000 quenched random configurations with $r_e = 1.5$ cm. A time step $\Delta t = 0$ in kilombo represents a single spatial configuration, or a single snapshot of the dynamics.

Our results show that for the small system size $N = 35$, the quenched kilobot configurations exhibit a slightly larger average degree than single snapshots of the kilombo dynamics simulations with $\Delta t = 0$. Again, this may be due to some bots being accumulated at the arena wall and not occupying space as effectively as they would do when randomly distributed, as shown in Chapter 7. This finite size effect reduces a bit the average connectivity of the kilobot network. However, when we increase the system size, the wall effect becomes negligible, and both the quenched and the $\Delta t = 0$ configurations yield approximately the same average degree. In all cases, increasing the exploratory time step and interaction radius leads to larger average degrees. This can be related with the average degree needed in a general network to

observe a giant component, which can be obtained from the celebrated Molloy and Reed (1995) criterion

$$\frac{\langle k^2 \rangle}{\langle k \rangle} > 2. \tag{9.2}$$

In the ideal case of the geometric network obtained for a uniform distribution of points, corresponding to zero size kilobots, the degree distribution can be shown to have the form of a homogeneous Poisson distribution (Ferrero and Gandino, 2011). For this distribution, we have $\langle k^2 \rangle = \langle k \rangle^2 + \langle k \rangle$, which, introduced into the formula of the Molloy and Reed criterion, predicts the position of the percolation threshold (signaling the presence of a giant component) at $\langle k \rangle = 1$. As we have seen in Figure 9.11, in the case of actual kilobots the degree distribution has also a Poisson-like form for different values of Δt and r_i. We can thus conclude that the dependence of the average degree on the interaction radius shown in Figure 9.12 is compatible, within this network interpretation, with a threshold radius of interaction that decreases with the time integration interval Δt.

9.3.3 Crowding effects in consensus reaching

In this section, we investigate the role of crowding effects in collective decision making in our nest-site selection model system. We also characterize the interplay between system size N, infrared communication distance or interacting radius r_i, and the radius R of the arena, for various conditions represented by the main model parameters, such as the interdependence parameter λ, and through several observables of interest.

We first describe crowding effects in terms of the agents' radius of interaction r_i by inspecting, for instance, the stationary values of dance frequencies $\langle f_i \rangle$ for different values of r_i. Large enough values of r_i should provide dance frequencies equivalent to mean-field results, whereas small r_i values, around a bot-diameter, would represent an ensemble of effectively independent bots. Remember that the communication radius is a fixed parameter equal to 7 cm in our experimental kilobot ensemble.

Taking advantage of the scalability properties of numerical simulations, we first

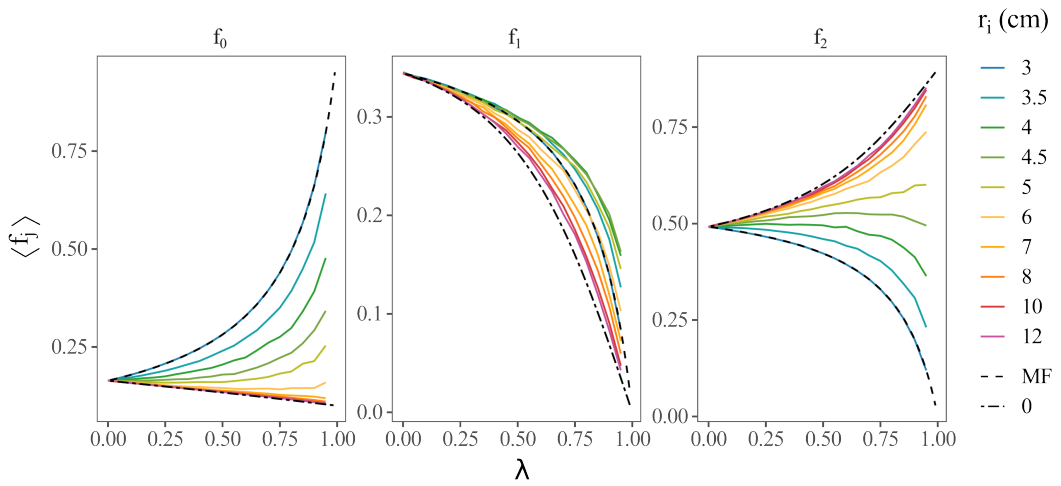


Figure 9.13: Stationary dance frequencies as a function of the interdependence parameter λ for different values of the communication radius r_i. For comparison, we also show the results of two limiting cases: the mean-field approximation (shown as dot-dashed lines), and the extreme case of isolated kilobots $r_i = 0$ (shown as dashed lines).

describe the model results for uncorrelated bot positions in the quenched approximation. In Figure 9.13, we plot the stationary values of $\langle f_i \rangle$ as a function of the interdependence parameter λ, and for fixed values of the interaction radius within the interval $3 \leq r_i \leq 12$ cm. The number of bots in these simulations is $N = 35$, and the radius of the arena is $R = 20$ cm, as in the kilobots experiments. In addition, we fix quality values q_1 and q_2 for both sites 1 and 2, which are equal to $q_1 = 7$ and $q_2 = 10$, and the values of the independent discovery probabilities $\pi_1 = \pi_2 = 0.3$, which are considered to be equal and, thus, correspond to the symmetric discovery scenario. For the sake of comparison, we also represent as dotted-dashed lines the expected stationary values in the mean field approximation, where every bot can exchange information with every other bot in the system, and the limiting case where bots remain isolated (shown as dashed lines). The latter limiting case would corresponds to an effective interacting radius equal to $r_i = 0$ in the simulations.

We see that asymptotically, at large enough interaction radius, the stationary frequencies converge towards mean field predictions (dotted-dashed lines). As we saw in the previous section, beyond an interaction radius approximately equal to $r_i^* \simeq 6.5 \pm 0.2$ we have a percolating communication network that enables the exchange of information among almost all bots in the system, and for this reason, the stationary values of the fraction of bots dancing for each state are very close to mean-field results. On the other hand, as the interaction radius decreases below

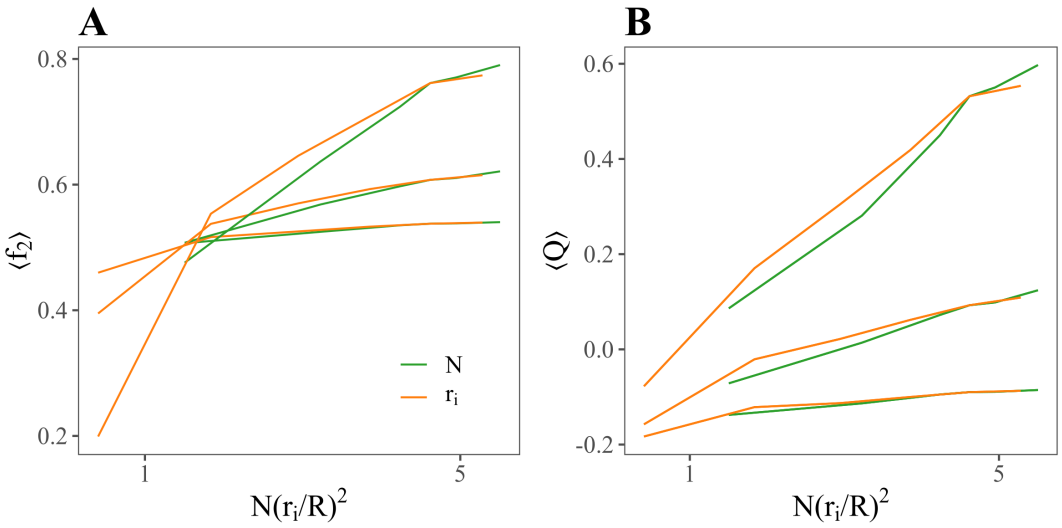


Figure 9.14: **A** Stationary dance frequency for the high quality site $\langle f_2 \rangle$, and **B** consensus parameter $\langle Q \rangle$ as a function of $N(r_i/R)^2$ in uncorrelated quenched configurations of bots, for three values of the interdependence parameter $\lambda = 0.3, 0.6, 0.9$.

the percolation threshold r_i^*, average values deviate substantially from mean-field predictions, specially so for higher values of the interdependence parameter λ, i.e. when communication matters the most. In the opposite limit, when $\lambda = 0$, the value of the interaction radius becomes an irrelevant variable. In this case, all curves match the $\lambda = 0$ analytic result of List et al., 2009 model that we discussed in the previous Chapter 8 (see Equations 8.13 and 8.14). When the interaction radius $r_i \leq 3$, due to excluded volume effects (each bot has an excluded radius equal to 1.5 cm), there are no bots within the intercommunication distance, and stationary values follow an effective zero interaction limit (as when $\lambda = 0$) but, in this case, with effective *a priori* discovery probabilities equal to $(1 - \lambda)\Pi_j$ in each case. This is also an analytical limiting result that one can find exactly from the definition of the model under scrutiny. It is represented with a dashed line in Figure 9.13.

Second, we investigate crowding effects by directly changing the density of bots in the arena. Within the same quenched approximation, we analyze the stationary behavior of the dancing frequency for the best-quality site and the average consensus parameter for different values of the number of bots N within the interval $5 \leq N \leq 100$ in the same arena of radius $R = 20$ cm. Our simulation results for different values of N with a fixed value of r_i, and for different values of r_i with a fixed number of bots N are represented together to characterize the scaling properties

of relevant observables, such as dancing frequencies and the consensus parameter, with these quantities. Our results confirm the basic assumption that for a homogeneous, uncorrelated, distribution of bots in the arena, the interaction area of each kilobot should scale linearly with the inverse of the number density of kilobots, or in other words, that r_i should be simply rescaled by a factor $R/N^{1/2}$ to obtain the same functional dependence of both quantities for different values of N, r_i, and R. In Figures 9.14, we represent the behavior of $\langle f_2 \rangle$ and $\langle Q \rangle$ as a function of $N * (r_i/R)^2$ obtained for different combinations of N and r_i in the quenched approximation. Results for varying numbers of bots are depicted with an orange continuous line, whereas results for varying interaction radii are depicted with a green continuous line. One can observe a very reasonable collapse of both curves, in the three conditions considered with interdependence parameter values equal to $\lambda = 0.3, 0.6, 0.9$ in the symmetric probability discovery scenario.

Finally, we analyze crowding effects in an ensemble of moving kilobots by means of kilombo simulations at different bot densities. Even though kilobots have a built-in interaction radius of approximately $r_i = 7$ cm, calibration results showed that the number of bots seen by physical kilobots was compatible with a lower interaction radius of $r_i = 5 \pm 2$ cm in kilombo simulations. Here, we fix this interaction radius, as well as the radius of the arena $R = 20$ cm, and change the density by changing the number of bots in the range $5 \leq N \leq 35$.

Figure 9.15A shows the probability distribution of bots detected by a focal kilobot i in a time step of $\Delta t = 8.24$ seconds (or 800 kilobot loop iterations), B_{i,Δ_t}, as the number density of bots increases by changing N from $N = 5$ until $N = 35$. More interestingly, in Figure 9.15B, we represent the progression of the mean ratio of type 1 and type 2 bots detected by a focal kilobot i during a Δt cycle, $f_{i,j}$, as N is increased, for the same three contrasting values of $\lambda = 0.3, 0.6, 0.9$ considered in the quenched approximation.

Mean values of $f_{i,j}$, and of the average stationary dance frequencies $\langle f_j \rangle$ increase gradually with N until they reach a plateau for a group of approximately 30 kilobots. This means that even for lower number densities than the one considered in the physical kilobot experiments (where $N = 35$), the system achieves global information about dancing frequencies. Again, this is a signature of the formation of a

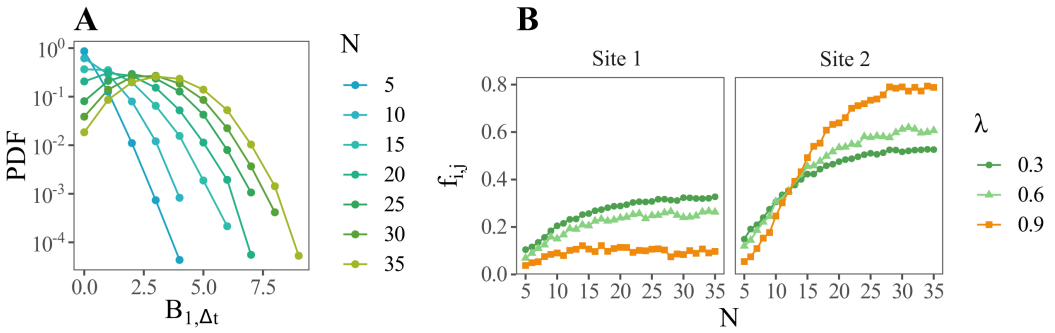


Figure 9.15: Bots detected during a time step Δt in kilombo simulations as a function of the number N of kilobots in the arena. **A**: Probability distribution, $B_{i,\Delta t}$, of bots detected by a focal kilobot i in a time step Δt. **B**: Ratio of type j bots, bots dancing for site j, detected by a focal kilobot i in a time step Δt, $f_{i,j}$, as a function of the number of kilobots in the arena.

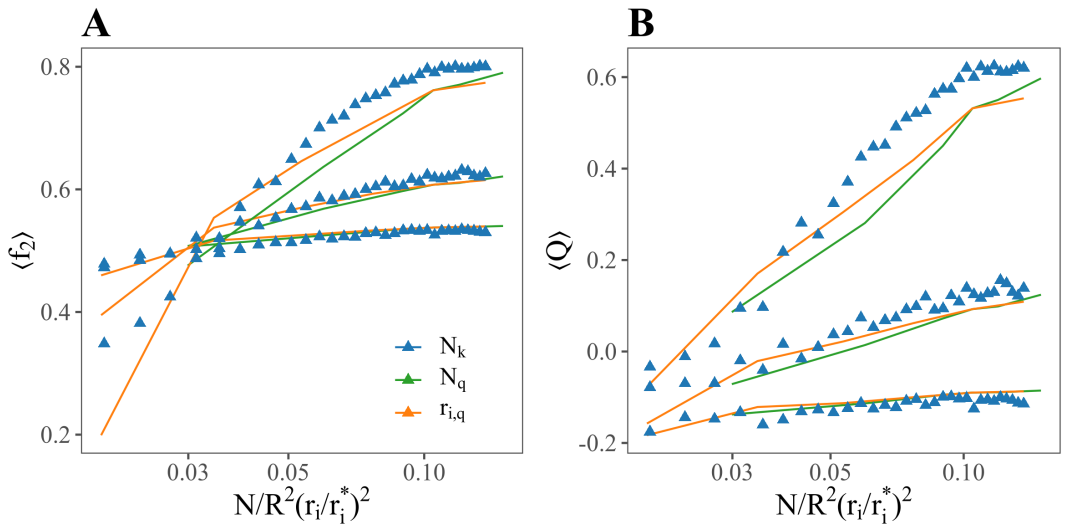


Figure 9.16: **A** Stationary dance frequency for the high quality site $\langle f_2 \rangle$, and **B** consensus parameter $\langle Q \rangle$ as a function of $N(r_i/R)^2$. Results are obtained from: (i) different ensemble sizes N and a fixed value of r_i (curve in orange), (ii) from different interaction radii r_i and a constant number of $N = 35$ bots (curve in green), both in the quenched position approximation; and (iii) from kilobots simulations in kilombo (triangular symbols), for three values of the interdependence parameter $\lambda = 0.3, 0.6, 0.9$. The interaction radius r_i in each case has been rescaled by the corresponding percolation threshold.

percolating communication network as a function of bot density for a fixed communication radius r_i. The movement of kilobots over the exploratory time window Δt, exchanging information within a radius r_i in an arena of radius R, enables the percolation of the communication network for number densities above a threshold value N^*, which facilitates the transmission of information throughout the system.

The average stationary values $\langle f_2 \rangle$ and $\langle Q \rangle$ obtained from kilombo simulations for different N values are also represented with triangular symbols in Figures 9.16 A and B.

The reasonable collapse of the kilombo simulation data with the quenched approximation results, orange and green curves, has been obtained after re-scaling the interaction radii by the corresponding percolation threshold r_i^* in each case. This feature corroborates our main starting hypothesis: the motion of kilobots over the exploring time window, together with its number density, are relevant variables that control the formation of a percolated communication network in the system, which facilitates the wide-spreading of information throughout the system, and determine the scaling properties of the most significant observables in collective decision making.

9.4 Closing remarks

In this study, we analyze how adding space and local interactions affects consensus reaching in the model proposed by List et al., 2009 of honeybees choosing a new site to build their nest. We chose kilobots as our study subjects since they are small swarm robots capable of performing collective tasks through exchanging information within an interaction radius.

In our version of the model, kilobots move randomly through a circular arena while exchanging information with nearby robots during a *sufficiently* long time step. We create an experimental setup that allows us to monitor the system at each time step, and also implement the model in a kilobots simulator, kilombo, to gain scalability and statistics. To better understand the impact of locality and the importance of kilobots' spatial mixing in collective decision-making, we compared our results with those obtained from a mean-field approximation and from quenched bots configurations. Our findings suggest that kilobots and quenched configurations achieve consensus levels as high as those obtained in a mean-field approximation in the low and medium communication regimes. However, only kilobots, through their mobility, achieve a similar level of consensus to the mean-field approximation in the high communication regime.

We investigate this further by using percolation theory and network methodologies and find that both the motion of kilobots and the density of the system play

a crucial role in forming a connected communication network. Through a percolated communication network, each bot can receive information beyond its immediate surroundings. The time kilobots spent exploring the environment results in an extended interaction radius, which reduces the bot density required to create a connected network in respect to a static system.

10 Conclusions and Perspectives

Collective animal behavior is a fascinating and complex phenomenon that involves the coordination of individuals in a group towards a common goal. Studying collective animal behavior offers insights into how individuals interact and exchange information within a group, with implications for a broad range of disciplines, including ecology, evolution, physics, and robotics. Moreover, it has also inspired the development of new algorithms and technologies. By developing agent-based or SPP models, researchers have been able to simulate collective animal behavior using simple local interaction rules and to create more efficient algorithms for decentralized decision-making that are applicable to network systems, robotics and optimization processes. In addition, technological advancements have allowed for the collection of vast amounts of high-precision behavioral data, leading to diverse technological solutions to address scenarios where animals need to be quantified.

Although the unifying theory for self-organization is still a challenge, self organizing groups share common principles that contribute to the emergence of collective behavior in animals. Group variability, positive and negative feedbacks, response thresholds, and effective leadership all play vital roles in discrete and continuum collective decision-making processes. Behavioral transitions observed in animal collectives also represent an esential adaptive property which implies a continuous adjustment of individuals' interactions to enable a swift response to unexpected stimuli, while also being robust enough to maintain the current behavior.

In this work, our main objective is to investigate the underlying mechanisms that give rise to complex collective behaviors in animals. Specifically, we concentrate on two types of collective behavior: collective motion and collective decision-making. By analyzing these phenomena, we aim to shed light on the ways in which animals communicate and self-organize.

Our investigation into collective motion focus on two different but interconnected behaviors in schooling fish. First, we analyze collective ordering dependence on swimming speed. To accomplish this, we take advantage of the natural speed variations observed in our study fish, which are primarily the result of the species' type of locomotion. Secondly, we narrow our attention to a specific type of directional rearrangement: the avalanche behaviors that are observed in our schooling fish. We explore the connection between these behaviors and effective leadership, seeking to better understand the relationship between them.

To analyze collective decision-making, we implement a model of honeybees nest site selection in physical swarming robots, which incorporates space and locality, as well as physical restrictions.

Our main goal gives rise to two additional objectives. On the one hand, in order to empirically investigate animal collective behavior, it is crucial to first establish appropriate experimental designs and setups. In this study, we place great emphasis on developing the necessary tools for recording, tracking, and processing our empirical data. Many of the resources that we use are custom made. This is a highly challenging and time-consuming undertaking that is essential for producing accurate and reliable results. On the other hand, we modify existing models of collective behavior to explain our empirical observations. The ability of these models to replicate observed collective behaviors increases their value and broadens their applicability. We also undertake a detailed exploration of the implemented models' parameters to gain a better understanding of their characteristics and behaviors, and we uncover several unreported properties of the models. Next, we describe the main findings of this work.

Our study of black neon tetra and the modification of the Vicsek model by including a periodic variable particles' speed sheds light on how collective order depends on the speed of individuals within a group. We found strong evidence of the correlation between swimming speed and collective ordering in fish collectives. This relationship implies that an increase in group speed is equivalent to increasing the alignment interaction strength and, thus, decreasing the effect of noise in fish matching orientations. This phase transition can be understood as a competition between noise and social interactions. From a biological perspective, a positive correlation

between speed and collective order could serve as an adaptive mechanism that ensures rapid information transfer in changing environmental conditions. For example, when fish feel threatened, they increase their swimming speed and ordering to collectively respond to attacks.

Through our exploration of the order parameter in the original Vicsek model we have uncovered a strong relationship between the system ordering and particles' speed. The fact that a disorder-to-order phase transition is observed when increasing particles' speed in a simple model like the Vicsek model contributes to a better understanding of the possible universality of phase transitions in moving groups.

Our study on behavioral cascades in our fish collective shows that scale-free avalanches can be observed in the rearrangement dynamics of swimming fish schools. The presence of effective leadership in fish schools can account for these avalanche behaviors, as some fish have an unusually large probability to initiate any avalanche in which they participate. This leadership can lead to sudden behavioral changes or provide specific information about the environment, such as the proximity of a wall. By modifying the classical Vicsek model to include a global leader, our study replicates the intermittent scale-free behavior observed in fish schools, with avalanche size and duration distributions displaying a self-similar scaling form. Our findings contribute to the understanding of the interplay between leadership and behavioral cascades in animal groups.

For our study of collective decision-making, we investigate the behavior of a decentralized decision-making model aimed at shedding light on the mechanisms used by honeybees to choose the best option for a new nest site. The model integrates independent discovery of potential sites and inter-individual communication to achieve consensus toward the best site, with site quality expressed through dance duration. The variance in dance times is a driving indicator for breaking system symmetry, as high-quality sites receive more "advertisement", generating more dancing activity for that site. The interdependence parameter represents the level of communication between agents. The level of communication needed to achieve consensus toward the high-quality site increases with the probability of the low-quality site being discovered and with its similarity from the high-quality site. Independence in this model acts as random noise that hampers the system's ability to reach

consensus, especially when sites are easy to discover. We analyze the feasibility of this model as a possible mechanism for honeybee nest-site selection by investigating the effects of more realistic interactions on consensus reaching. Our results demonstrate that physical kilobots, through their mobility and interaction with the local environment, form a percolating communication network that allows them to effectively sample the state of the system and transmit information to achieve collective decision-making comparable to mean-field interactions. The impact of communication on individual decision making is reduced with low interdependence, but becomes more relevant for consensus reaching as interdependence increases. Within the high interdependence regime, the mobility of agents, such as kilobots, is needed for reaching high consensus values as those obtained in mean-field simulations. The dynamic nature in our swarming robots implementation of the model induces an extension of the interaction radius in a static scenario. Our study provides insights about the mechanisms underlying collective decision-making, specially regarding individuals' connectivity for efficient information transfer. Our findings demonstrate how honeybees can achieve consensus on the best available site for their nest solely through local dynamical interactions. Our results increase the robustness of a model that was previously only implemented in unrealistic mean-field approximations.

As mentioned, working with empirical data can be challenging due to various reasons. Firstly, collecting data can be time-consuming and resource-intensive. Secondly, analyzing empirical data can be complex, especially when attempting to draw causal inferences from the data. Moreover, having a sufficient amount of data can also be a challenge, as researchers need to ensure that they have enough data points to achieve statistical significance and to make meaningful conclusions. Although our data provides valuable findings, there may be additional interesting questions that could not be fully addressed with the current data.

One area of interest in the study of fish schools is gaining a more detailed understanding of how the boundaries of the school's environment affect their behavior. While variations in speed and sudden directional rearrangments are largely the result of behavioral synchronization within the collective, the confinement of the fish to a tank makes it important to separate the effects of the tank from those of the

group interactions to fully understand the adaptive behavior of the school. In this sense, empirical analysis of schooling fish in tanks of different sizes and with varying group sizes would be necessary. This could provide valuable insights into the ways in which fish schools adapt to environmental barriers and navigate through their surroundings.

In future research, exploring the role of individual variation in collective behavior could provide valuable insights into how group composition influences collective behavior and how individuals interact within a group. For instance, investigating the effect of leader switching strategies on the preservation of the observed scale-free nature of avalanches under different conditions could be a promising direction. Moreover, further studies could investigate the dynamical networks of larger fish schools and additional temporal series, enabling the examination of the potential preferences of individual fish to stay close to conspecifics.

Introducing individual variation in collective motion and decision-making models can lead to a more comprehensive understanding of the diverse strategies and adaptive behaviors that underlie animal collectives. For example, by incorporating bees that are more experienced and carry more weight in the decision-making process or individuals that can learn and change their "social status" within the collective. More realistic models incorporating social networks, cognitive limitations, and stress effects on decision-making can predict how animal groups make decisions and explore how these decisions may change in response to different environmental conditions.

With this work, we contribute to the understanding of the ways in which animals can communicate and self-organize. We accomplish this through the development of appropriate experimental setups and modifications of existing models of collective behavior. Through our research, we have uncovered several new properties of these existing models and identified relationships between collective ordering and individuals' speed, the interplay between leadership and behavioral cascades and the effect of dynamical local interactions in consensus reaching. Our research findings aim to inspire and encourage further exploration in this field. Our hope is to contribute to the development of more comprehensive models and to expand the knowledge regarding the universality of collective behavior.

Publications

1. Múgica, J., Torrents, J., Cristín, J., Puy, A., Miguel, M. C., & Pastor-Satorras, R. (2022). Scale-free behavioral cascades and effective leadership in schooling fish. *Scientific reports*, 12(1), 10783.

2. Múgica J., Miguel, M.C., & Pastor-Satorras, R. (2023). Quasiperiodic collective polar ordering in fish schools and its dependence on swimming speed: An experimental and theoretical comparison. Submitted to *PLOS Computational Biology*.

3. Mars D., Múgica J., Ferrero E., & Miguel, M.C. (2023) Honey bee-like collective decision-making in a kilobot swarm. In preparation.

Bibliography

Abbot, Patrick (2022). "Defense in social insects: diversity, division of labor, and evolution". In: *Annual Review of Entomology* 67, pp. 407–436.

Aldana, M, H Larralde, and B Vázquez (2009). "On the emergence of collective order in swarming systems: a recent debate". In: *International Journal of Modern Physics B* 23.18, pp. 3661–3685.

Altshuler, E. and T. H. Johansen (2004). "Colloquium: Experiments in vortex avalanches". In: *Rev. Mod. Phys.* 76 (2), pp. 471–487. DOI: 10.1103/RevModPhys.76.471. URL: https://link.aps.org/doi/10.1103/RevModPhys.76.471.

Alvarez-Buylla, Elena R et al. (2008). "Gene regulatory network models: A dynamic and integrative approach to development". In: *Practical Systems Biology*, pp. 127–154.

Ansari, Shamshad (2020). "Techniques of Image Processing". In: *Building Computer Vision Applications Using Artificial Neural Networks*. Springer, pp. 27–94.

Ariel, Gil and Amir Ayali (2015). "Locust Collective Motion and Its Modeling". In: *PLOS Computational Biology* 11.12. Ed. by Frederick R. Adler, e1004522. ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1004522. URL: http://dx.plos.org/10.1371/journal.pcbi.1004522.

Arvin, Farshad et al. (2011). "Imitation of honeybee aggregation with collective behavior of swarm robots". In: *International Journal of Computational Intelligence Systems* 4.4, pp. 739–748.

Ashraf, I et al. (2016). "Synchronization and collective swimming patterns in fish (Hemigrammus bleheri)". In: *Journal of the Royal Society Interface* 13.123, p. 20160734.

Attanasi, Alessandro et al. (2014). "Finite-size scaling as a way to probe near-criticality in natural swarms". In: *Physical review letters* 113.23, p. 238102.

Attanasi, Alessandro et al. (2015). "Emergence of collective changes in travel direction of starling flocks from individual birds' fluctuations". In: *Journal of The Royal Society Interface* 12.108, p. 20150319.

Aubé, François and Robert Shield (2004). "Modeling the Effect of Leadership on Crowd Flow Dynamics". In: *Lect. Notes Comput. Sci.* Vol. 3305. Springer Berlin / Heidelberg, pp. 601–611. ISBN: 978-3-540-23596-5. DOI: 10.1007/978-3-540-30479-1_62. URL: http://dx.doi.org/10.1007/978-3-540-30479-1{_}62http://link.springer.com/10.1007/978-3-540-30479-1{_}62.

Baddeley, Adrian, Imre Bárány, and Rolf Schneider (2007). "Spatial point processes and their applications". In: *Stochastic Geometry: Lectures Given at the CIME Summer School Held in Martina Franca, Italy, September 13–18, 2004*, pp. 1–75.

Baglietto, Gabriel and Ezequiel V. Albano (2009a). "Computer simulations of the collective displacement of self-propelled agents". In: *Computer Physics Communications* 180.4, pp. 527–531. ISSN: 0010-4655. DOI: 10.1016/j.cpc.2008.12.026.

— (2009b). "Computer simulations of the collective displacement of self-propelled agents". In: *Computer Physics Communications* 180.4, pp. 527–531. ISSN: 0010-4655. DOI: 10.1016/j.cpc.2008.12.026.

Balda, Russell P and Gary C Bateman (1971). "Flocking and annual cycle of the pinon jay, Gymnorhinus cyanocephalus". In: *Condor*, pp. 287–302.

Barber, Iain and Hazel A Wright (2001). "How strong are familiarity preferences in shoaling fish?" In: *Animal Behaviour* 61.5, pp. 975–979.

Barrat, Alain, Marc Barthélemy, and Alessandro Vespignani (2008). *Dynamical Processes on Complex Networks*. Cambridge: Cambridge University Press.

Barthélemy, Marc (2011). "Spatial networks". In: *Physics Reports* 499.1, pp. 1–101.

Beckers, Ralph, Jean-Louis Deneubourg, and Simon Goss (1992). "Trails and U-turns in the selection of a path by the ant Lasius niger". In: *Journal of theoretical biology* 159.4, pp. 397–415.

— (1993). "Modulation of trail laying in the ant Lasius niger (Hymenoptera: Formicidae) and its role in the collective selection of a food source". In: *Journal of Insect Behavior* 6, pp. 751–759.

Beekman, Madeleine and Benjamin P Oldroyd (2018). "Different bees, different needs: how nest-site requirements have shaped the decision-making processes in homeless honeybees (Apis spp.)" In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 373.1746, p. 20170010.

Beekman, Madeleine, David JT Sumpter, and Francis LW Ratnieks (2001). "Phase transition between disordered and ordered foraging in Pharaoh's ants". In: *Proceedings of the National Academy of Sciences* 98.17, pp. 9703–9706.

Biro, Dora, Takao Sasaki, and Steven J Portugal (2016). "Bringing a time–depth perspective to collective animal behaviour". In: *Trends in ecology & evolution* 31.7, pp. 550–562.

Bode, Nikolai W.F., A. Jamie Wood, and Daniel W. Franks (2011). "The impact of social networks on animal collective motion". In: *Anim. Behav.* 82.1, pp. 29–38. ISSN: 00033472. DOI: 10.1016/j.anbehav.2011.04.011. URL: http://dx.doi.org/10.1016/j.anbehav.2011.04.011http://linkinghub.elsevier.com/retrieve/pii/S0003347211001539.

Bonabeau, Eric, Guy Theraulaz, and Jean-Louis Deneubourg (1998). "Fixed response thresholds and the regulation of division of labor in insect societies". In: *Bulletin of Mathematical Biology* 60.4, pp. 753–807.

Boolchand, P et al. (2005). "Self-organization and the physics of glassy networks". In: *Philosophical Magazine* 85.32, pp. 3823–3838.

Borrel, Vincent et al. (2008). "Simps: Using sociology for personal mobility". In: *IEEE/ACM transactions on networking* 17.3, pp. 831–842.

Bose, Thomas, Andreagiovanni Reina, and James AR Marshall (2017). "Collective decision-making". In: *Current opinion in behavioral sciences* 16, pp. 30–34.

Bownik, Adam and Donald Wlodkowic (2021). "Applications of advanced neurobehavioral analysis strategies in aquatic ecotoxicology". In: *Science of the Total Environment* 772, p. 145577.

Branson, Kristin et al. (2009). "High-throughput ethomics in large groups of Drosophila". In: *Nature methods* 6.6, pp. 451–457.

Britton, Nick F et al. (2002). "Deciding on a new home: how do honeybees agree?" In: *Proceedings of the Royal Society of London. Series B: Biological Sciences* 269.1498, pp. 1383–1388.

Buhl, J. et al. (2006). "From Disorder to Order in Marching Locusts". In: *Science* 312.5778, pp. 1402–1406. ISSN: 0036-8075. DOI: 10.1126/science.1125142. URL: http://www.sciencemag.org/cgi/doi/10.1126/science.1128057http://www.sciencemag.org/cgi/doi/10.1126/science.1125142.

Cachat, Jonathan et al. (2011). "Three-dimensional neurophenotyping of adult zebrafish behavior". In: *PloS one* 6.3, e17597.

Calovi, Daniel S et al. (2014). "Swarming, schooling, milling: phase diagram of a data-driven fish school model". In: *New journal of Physics* 16.1, p. 015026.

Calovi, Daniel S et al. (2015). "Collective response to perturbations in a data-driven fish school model". In: *Journal of The Royal Society Interface* 12.104, p. 20141362.

Calovi, Daniel S. et al. (Jan. 2018). "Disentangling and modeling interactions in fish with burst-and-coast swimming reveal distinct alignment and attraction behaviors". In: *PLOS Computational Biology* 14.1, pp. 1–28. DOI: 10.1371/journal.pcbi.1005933. URL: https://doi.org/10.1371/journal.pcbi.1005933.

Camazine, Scott et al. (2001a). *Self-organization in biological systems*. Princeton Studies in Complexity. Princeton, New Jersey: Princeton University Press.

— (2001b). "Self-organization in biological systems". In: *Self-Organization in Biological Systems*. Princeton university press.

Cambui, Dorilson S and Alexandre Rosas (2018). "Transitions induced by speed in self-propelled particles system with attractive interactions". In: *Physica A: Statistical Mechanics and its Applications* 498, pp. 96–101.

Camperi, Marcelo et al. (2012). "Spatially balanced topological interaction grants optimal cohesion in flocking models". In: *Interface focus* 2.6, pp. 715–725.

Campo, Alexandre et al. (2010). "Artificial pheromone for path selection by a foraging swarm of robots". In: *Biological cybernetics* 103, pp. 339–352.

Cardy, J. L., ed. (1988). *Finite Size Scaling*. Vol. 2. Current Physics-Sources and Comments. Amsterdam: North Holland.

Cardy, John (1996). *Scaling and Renormalization in Statistical Physics*. Cambridge Lecture Notes in Physics. Cambridge University Press. DOI: 10.1017/CBO9781316036440.

Carvalho, Fernando R and Francisco Langeani (2013). "Hyphessobrycon uaiso: new characid fish from the rio Grande, upper rio Paraná basin, Minas Gerais State

(Ostariophysi: Characidae), with a brief comment about some types of Hyphessobrycon". In: *Neotropical Ichthyology* 11, pp. 525–536.

Castellano, Claudio, Santo Fortunato, and Vittorio Loreto (2009). "Statistical physics of social dynamics". In: *Reviews of modern physics* 81.2, p. 591.

Chen, D. et al. (2016). "Switching hierarchical leadership mechanism in homing flight of pigeon flocks". In: *Europhys. Lett.* 114, p. 60008.

Clark, Philip J and Francis C Evans (1954). "Distance to nearest neighbor as a measure of spatial relationships in populations". In: *Ecology* 35.4, pp. 445–453.

Clusella, Pau and Romualdo Pastor-Satorras (2021). "Phase transitions on a class of generalized Vicsek-like models of collective motion". In: *Chaos: An Interdisciplinary Journal of Nonlinear Science* 31.4, p. 043116.

Conradt, Larissa (2012). "Models in animal collective decision-making: information uncertainty and conflicting preferences". In: *Interface focus* 2.2, pp. 226–240.

Conradt, Larissa and Timothy J Roper (2000). "Activity synchrony and social cohesion: a fission-fusion model". In: *Proceedings of the Royal Society of London. Series B: Biological Sciences* 267.1458, pp. 2213–2218.

Cont, Rama and Jean-Philipe Bouchaud (2000). "Herd behavior and aggregate fluctuations in financial markets". In: *Macroeconomic dynamics* 4.2, pp. 170–196.

Couzin, I D et al. (2005). "Effective leadership and decision-making in animal groups on the move". In: *Nature* 433.7025, pp. 513–516. ISSN: 1476-4687. DOI: 10.1038/nature03236. URL: http://www.ncbi.nlm.nih.gov/pubmed/15690039.

Couzin, I. D. et al. (2011). "Uninformed Individuals Promote Democratic Consensus in Animal Groups". In: *Science* 334.6062, pp. 1578–1580. ISSN: 0036-8075. DOI: 10.1126/science.1210280.

Couzin, Iain (2007). "Collective minds". In: *Nature* 445.7129, pp. 715–715.

Couzin, Iain D, Jens Krause, et al. (2003). "Self-organization and collective behavior in vertebrates". In: *Advances in the Study of Behavior* 32.1, pp. 10–1016.

Couzin, ID et al. (2002). "Collective Memory and Spatial Sorting in Animal Groups". In: *Journal of Theoretical Biology* 218.1, pp. 1–11.

Croft, Darren P et al. (2004). "Intra-sexual preferences for familiar fish in male guppies". In: *Journal of Fish Biology* 64.1, pp. 279–283.

Czirók, András, H. Eugene Stanley, and Tamás Vicsek (1997). "Spontaneously ordered motion of self-propelled particles". In: *Journal of Physics A: Mathematical and General* 30.5, pp. 1375–1385. ISSN: 0305-4470. DOI: 10.1088/0305-4470/30/5/009.

Daniel Kissling, W, David E Pattemore, and Melanie Hagen (2014). "Challenges and prospects in the telemetry of insects". In: *Biological Reviews* 89.3, pp. 511–530.

De Croon, GCHE et al. (2022). "Insect-inspired AI for autonomous robots". In: *Science Robotics* 7.67, eabl6334.

De Menech, M., A. L. Stella, and C. Tebaldi (1998). "Rare events and breakdown of simple scaling in the Abelian sandpile model". In: *Phys. Rev. E* 58 (3), R2677–R2680. DOI: 10.1103/PhysRevE.58.R2677. URL: https://link.aps.org/doi/10.1103/PhysRevE.58.R2677.

Deffuant, Guillaume et al. (2000). "Mixing beliefs among interacting agents". In: *Advances in Complex Systems* 3.01n04, pp. 87–98.

Deisboeck, Thomas S and Iain D Couzin (2009). "Collective behavior in cancer cell populations". In: *Bioessays* 31.2, pp. 190–197.

Delcourt, Johann and Pascal Poncin (2012). "Shoals and schools: back to the heuristic definitions and quantitative references". In: *Reviews in Fish Biology and Fisheries* 22.3, pp. 595–619.

Dell, Anthony I et al. (2014). "Automated image-based tracking and its application in ecology". In: *Trends in ecology & evolution* 29.7, pp. 417–428.

Deutsch, Andreas, Guy Theraulaz, and Tamas Vicsek (2012). "Collective motion in biological systems". In: *Interface Focus* 2.6, pp. 689–692. DOI: 10.1098/rsfs.2012.0048. eprint: https://royalsocietypublishing.org/doi/pdf/10.1098/rsfs.2012.0048. URL: https://royalsocietypublishing.org/doi/abs/10.1098/rsfs.2012.0048.

Dimidov, Cristina, Giuseppe Oriolo, and Vito Trianni (2016). "Random walks in swarm robotics: an experiment with kilobots". In: *International conference on swarm intelligence*. Springer, pp. 185–196.

Dolado, Ruth et al. (2015). "A method for resolving occlusions when multitracking individuals in a shoal". In: *Behavior research methods* 47.4, pp. 1032–1043.

Doran, Carolina et al. (2022). "Fish waves as emergent collective antipredator behavior". In: *Current Biology* 32.3, pp. 708–714.

Dorigo, Marco, Guy Theraulaz, and Vito Trianni (2021). "Swarm robotics: Past, present, and future [point of view]". In: *Proceedings of the IEEE* 109.7, pp. 1152–1165.

Dussutour, Audrey et al. (2008). "Individual differences influence collective behaviour in social caterpillars". In: *Animal Behaviour* 76.1, pp. 5–16.

Dutil, J.-D. et al. (2007). "Burst and coast use, swimming performance and metabolism of Atlantic cod Gadus morhua in sub-lethal hypoxic conditions". In: *Journal of Fish Biology* 71.2, pp. 363–375. DOI: https://doi.org/10.1111/j.1095-8649.2007.01487.x. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1095-8649.2007.01487.x. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1095-8649.2007.01487.x.

Dyer, Fred C (2002). "The biology of the dance language". In: *Annual review of entomology* 47.1, pp. 917–949.

Dyer, John RG et al. (2009). "Leadership, consensus decision making and collective behaviour in humans". In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 364.1518, pp. 781–789.

Fernando, Prithiviraj and Russell Lande (2000). "Molecular genetic and behavioral analysis of social organization in the Asian elephant (Elephas maximus)". In: *Behavioral Ecology and Sociobiology* 48.1, pp. 84–91.

Ferrero, Renato and Filippo Gandino (2011). "Degree Distribution of Unit Disk Graphs with Uniformly Deployed Nodes on a Rectangular Surface". In: *2011 International Conference on Broadband and Wireless Computing, Communication and Applications*, pp. 255–262. DOI: 10.1109/BWCCA.2011.38.

Fish, Frank (Jan. 2010a). "Swimming Strategies for Energy Economy". In: *Fish Locomotion: An Eco-ethological Perspective*. DOI: 10.1201/b10190-5.

Fish, Frank E (2010b). "Swimming strategies for energy economy". In: *Fish swimming: an etho-ecological perspective* 90.

Fish, Frank E, Jennifer F Fegely, and Cindy J Xanthopoulos (1991). "Burst-and-coast swimming in schooling fish (Notemigonus crysoleucas) with implications for energy economy". In: *Comparative Biochemistry and Physiology Part A: Physiology* 100.3, pp. 633–637. ISSN: 0300-9629. DOI: https://doi.org/10.1016/0300-

9629(91)90382-M. URL: https://www.sciencedirect.com/science/article/pii/030096299190382M.

Fisher, D. S. (1998). "Collective Transport in Random Media: From Superconductors to Earthquakes". In: *Physics Reports* 301, pp. 113–150. ISSN: 03701573. DOI: 10.1016/S0370-1573(98)00008-8. arXiv: 9711179 [cond-mat]. URL: http://arxiv.org/abs/cond-mat/9711179.

Flack, Andrea et al. (2012). "What are leaders made of? The role of individual experience in determining leader-follower relations in homing pigeons". In: *Animal Behaviour* 83.3, pp. 703–709. ISSN: 00033472. DOI: 10.1016/j.anbehav.2011.12.018.

Fornberg, Bengt (1988). "Generation of finite difference formulas on arbitrarily spaced grids". In: *Mathematics of Computation* 51.184, pp. 699–699. ISSN: 0025-5718. DOI: 10.1090/S0025-5718-1988-0935077-0.

Foster, WA and JE Treherne (1981). "Evidence for the dilution effect in the selfish herd from fish predation on a marine insect". In: *Nature* 293.5832, pp. 466–467.

Franks, Nigel R et al. (2009). "Speed versus accuracy in decision-making ants: expediting politics and policy implementation". In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 364.1518, pp. 845–852.

Galam, Serge (2002). "Minority opinion spreading in random geometry". In: *The European Physical Journal B-Condensed Matter and Complex Systems* 25, pp. 403–406.

— (2005). "Heterogeneous beliefs, segregation, and extremism in the making of public opinions". In: *Physical Review E* 71.4, p. 046123.

Galla, Tobias (2010). "Independence and interdependence in the nest-site choice by honeybee swarms: agent-based models, analytical approaches and pattern formation". In: *Journal of Theoretical Biology* 262.1, pp. 186–196.

Gallois, Benjamin and Raphaël Candelier (2021). "FastTrack: An open-source software for tracking varying numbers of deformable objects". In: *PLOS Computational Biology* 17.2, e1008697.

García, Gloria Bueno et al. (2015). *Learning image processing with OpenCV*. Packt Publishing Birmingham.

Garnier, Simon (2011). "From ants to robots and back: How robotics can contribute to the study of collective animal behavior". In: *Bio-inspired self-organizing robotic systems*, pp. 105–120.

Garnier, Simon et al. (2008). "The embodiment of cockroach aggregation behavior in a group of micro-robots". In: *Artificial life* 14.4, pp. 387–408.

Garnier, Simon et al. (2009). "Self-organized aggregation triggers collective decision making in a group of cockroach-like robots". In: *Adaptive Behavior* 17.2, pp. 109–133.

Gauci, Melvin, Radhika Nagpal, and Michael Rubenstein (2018). "Programmable self-disassembly for shape formation in large-scale robot collectives". In: *Distributed Autonomous Robotic Systems: The 13th International Symposium*. Springer, pp. 573–586.

Gautrais, Jacques et al. (2012). "Deciphering interactions in moving animal groups". In: *Plos computational biology* 8.9, e1002678.

George, Ginu et al. (2018). "A survey on various median filtering techniques for removal of impulse noise from digital image". In: *2018 Conference on Emerging Devices and Smart Systems (ICEDSS)*. IEEE, pp. 235–238.

Gerlotto, François, Stratis Georgakarakos, and Peter K Eriksen (2000). "The application of multibeam sonar technology for quantitative estimates of fish density in shallow water acoustic surveys". In: *Aquatic Living Resources* 13.5, pp. 385–393.

Giardina, Irene (2008). "Collective behavior in animal groups: theoretical models and empirical studies". In: *HFSP journal* 2.4, pp. 205–219.

Gimeno, Elisabet et al. (2016). "Differences in shoaling behavior in two species of freshwater fish (Danio rerio and Hyphessobrycon herbertaxelrodi)." In: *Journal of Comparative Psychology* 130.4, p. 358.

Ginelli, Francesco (2016). "The physics of the Vicsek model". In: *The European Physical Journal Special Topics* 225, pp. 2099–2117.

Ginelli, Francesco et al. (2015). "Intermittent collective dynamics emerge from conflicting imperatives in sheep herds". In: *Proceedings of the National Academy of Sciences* 112.41, pp. 12729–12734. ISSN: 0027-8424. DOI: 10.1073/pnas.1503749112. URL: http://www.pnas.org/lookup/doi/10.1073/pnas.1503749112.

Gollapudi, Sunila (2019). "OpenCV with Python". In: *Learn Computer Vision Using OpenCV*. Springer, pp. 31–50.

Grégoire, Guillaume and Hugues Chaté (2004). "Onset of collective and cohesive motion". In: *Physical review letters* 92.2, p. 025702.

Griffiths, Siân W and Anne E Magurran (1999). "Schooling decisions in guppies (Poecilia reticulata) are based on familiarity rather than kin recognition by phenotype matching". In: *Behavioral Ecology and Sociobiology* 45.6, pp. 437–443.

Grünbaum, Daniel (2012). "A spatially explicit Bayesian framework for cognitive schooling behaviours". In: *Interface focus* 2.6, pp. 738–745.

Grüter, Christoph and Walter M Farina (2009). "The honeybee waggle dance: can we follow the steps?" In: *Trends in ecology & evolution* 24.5, pp. 242–247.

Halloy, José et al. (2007). "Social integration of robots into groups of cockroaches to control self-organized choices". In: *Science* 318.5853, pp. 1155–1158.

Harcourt, Alexander H and Kelly J Stewart (1994). "Gorillas' vocalizations during rest periods: signals of impending departure?" In: *Behaviour* 130.1-2, pp. 29–40.

Harmer, Aaron MT and Daniel B Thomas (2019). "pathtrackr: An r package for video tracking and analysing animal movement". In: *Methods in Ecology and Evolution* 10.8, pp. 1196–1202.

Helbing, Dirk (2012). *Social self-organization: Agent-based simulations and experiments to study emergent social behavior*. Springer.

Herbert-Read, James E (2016). "Understanding how animal groups achieve coordinated movement". In: *Journal of Experimental Biology* 219.19, pp. 2971–2983.

Herbert-Read, James E. et al. (2011). "Inferring the rules of interaction of shoaling fish". In: *Proceedings of the National Academy of Sciences* 108.46, pp. 18726–18731. ISSN: 0027-8424. DOI: 10.1073/pnas.1109355108.

Herbert-Read, James E et al. (2015). "Initiation and spread of escape waves within animal groups". In: *Royal Society open science* 2.4, p. 140355.

Hidalgo, Jorge et al. (2014). "Information-based fitness and the emergence of criticality in living systems". In: *Proceedings of the National Academy of Sciences* 111.28, pp. 10095–10100.

Hoeffding, Wassily (1956). "On the Distribution of the Number of Successes in Independent Trials". In: *The Annals of Mathematical Statistics* 27.3, pp. 713 –721. DOI: 10.1214/aoms/1177728178. URL: https://doi.org/10.1214/aoms/1177728178.

Holley, Richard A and Thomas M Liggett (1975). "Ergodic theorems for weakly interacting infinite systems and the voter model". In: *The annals of probability*, pp. 643–663.

Huth, Andreas and Christian Wissel (1992). "The simulation of the movement of fish schools". In: *Journal of theoretical biology* 156.3, pp. 365–385.

Jakobi, Nick (1997). "Evolutionary robotics and the radical envelope-of-noise hypothesis". In: *Adaptive behavior* 6.2, pp. 325–368.

Jansson, Fredrik et al. (2015). "Kilombo: a Kilobot simulator to enable effective research in swarm robotics". In: *arXiv preprint arXiv:1511.04285*.

Jeanson, Raphael et al. (2005). "Self-organized aggregation in cockroaches". In: *Animal behaviour* 69.1, pp. 169–180.

Judd, Timothy M (1995). "The Waggle Dance of the Honey Bee: Which Bees Following a Dancer Successfully Acquire the". In: *Journal of Insect Behavior* 8.3.

Karn, Phil et al. (1990). "MACA-a new channel access method for packet radio". In: *ARRL/CRRL Amateur radio 9th computer networking conference*. Vol. 140. London, Canada, pp. 134–140.

Kastberger, Gerald, Evelyn Schmelzer, and Ilse Kranner (2008). "Social waves in giant honeybees repel hornets". In: *PLoS One* 3.9, e3141.

Katz, Yael et al. (2011). "Inferring the structure and dynamics of interactions in schooling fish". In: *Proceedings of the National Academy of Sciences* 108.46, pp. 18720–18725. ISSN: 0027-8424. DOI: 10.1073/pnas.1107583108. eprint: https://www.pnas.org/content/108/46/18720.full.pdf. URL: https://www.pnas.org/content/108/46/18720.

Kawamura, Hikaru et al. (2012). "Statistical physics of fracture, friction, and earthquakes". In: *Rev. Mod. Phys.* 84 (2), pp. 839–884. DOI: 10.1103/RevModPhys.84.839. URL: https://link.aps.org/doi/10.1103/RevModPhys.84.839.

Kilolib (n.d.). https://kilobotics.com/docs/index.html. Accessed: 2022-11-05.

Klamser, Pascal P and Pawel Romanczuk (2021). "Collective predator evasion: Putting the criticality hypothesis to the test". In: *PLoS Computational Biology* 17.3, e1008832.

Korsmeyer, Keith E, John Fleng Steffensen, and Jannik Herskin (2002). "Energetics of median and paired fin swimming, body and caudal fin swimming, and gait transition in parrotfish (Scarus schlegeli) and triggerfish (Rhinecanthus aculeatus)". In: *Journal of experimental biology* 205.9, pp. 1253–1263.

Krakauer, David C (1995). "Groups confuse predators by exploiting perceptual bottlenecks: a connectionist model of the confusion effect". In: *Behavioral Ecology and Sociobiology* 36.6, pp. 421–429.

Kramer, Boris H, Inon Scharf, and Susanne Foitzik (2014). "The role of per-capita productivity in the evolution of small colony sizes in ants". In: *Behavioral Ecology and Sociobiology* 68, pp. 41–53.

Krause, Jens, David Lusseau, and Richard James (2009). "Animal social networks: an introduction". In: *Behavioral Ecology and Sociobiology* 63.7, pp. 967–973.

Krause, Jens, Alan FT Winfield, and Jean-Louis Deneubourg (2011). "Interactive robots in experimental biology". In: *Trends in ecology & evolution* 26.7, pp. 369–375.

Krause, Mark A, Karen L Hollis, and Mauricio R Papini (2022). "Perspectives on the Evolution of Learning and Memory Mechanisms". In: *Evolution of Learning and Memory Mechanisms*, p. 1.

Krieger, Michael JB, Jean-Bernard Billeter, and Laurent Keller (2000). "Ant-like task allocation and recruitment in cooperative robots". In: *Nature* 406.6799, pp. 992–995.

Kuramoto, Yoshiki (1975). "Self-entrainment of a population of coupled non-linear oscillators". In: *International Symposium on Mathematical Problems in Theoretical Physics: January 23–29, 1975, Kyoto University, Kyoto/Japan*. Springer, pp. 420–422.

Kyriakopoulos, Nikos, Francesco Ginelli, and John Toner (2016). "Leading birds by their beaks: the response of flocks to external perturbations". In: *New Journal of Physics* 18.7, p. 73039. DOI: 10.1088/1367-2630/18/7/073039. URL: https://doi.org/10.1088{\%}2F1367-2630{\%}2F18{\%}2F7{\%}2F073039.

Laurson, Lasse and Mikko J. Alava (2006). "1/f noise and avalanche scaling in plastic deformation". In: *Phys. Rev. E* 74.066106.

Laurson, Lasse, Xavier Illa, and Mikko J. Alava (2009). "The effect of thresholding on temporal avalanche statistics". In: *J. Stat. Mech.: Theo. Exp.* P01019.P01019.

Leadbeater, Ellouise and Lars Chittka (2007). "Social learning in insects—from miniature brains to consensus building". In: *Current biology* 17.16, R703–R713.

Lee-Jenkins, Stacey SY and Jean-Guy J Godin (2013). "Concurrent effects of familiarity and kinship on social affiliations in convict cichlid (Amatitlania siquia) young". In: *Behaviour* 150.8, pp. 895–919.

Lemasson, B. H., J. J. Anderson, and R. A. Goodwin (2013). "Motion-guided attention promotes adaptive communications during social navigation". In: *Proceedings of the Royal Society B: Biological Sciences* 280.1754, p. 20122003. DOI: 10.1098/rspb.2012.2003. eprint: https://royalsocietypublishing.org/doi/pdf/10.1098/rspb.2012.2003. URL: https://royalsocietypublishing.org/doi/abs/10.1098/rspb.2012.2003.

Li, G et al. (2020). "Burst-and-coast swimmers optimize gait by adapting unique intrinsic cycle". In: *arXiv preprint arXiv:2002.09176*.

Li, Gen et al. (2021). "Burst-and-coast swimmers optimize gait by adapting unique intrinsic cycle". In: *Communications biology* 4.1, pp. 1–7.

Liao, James C et al. (2003). "Fish exploiting vortices decrease muscle activity". In: *Science* 302.5650, pp. 1566–1569.

Lima, Flavio CT, Daniel P Coutinho, Wolmar B Wosiacki, et al. (2014). "A new Hyphessobrycon (Ostariophysi: Characiformes: Characidae) from the middle Amazon basin, Brazil". In: *Zootaxa* 3872.2, pp. 167–179.

List, Christian, Christian Elsholtz, and Thomas D Seeley (2009). "Independence and interdependence in collective decision making: an agent-based model of nest-site choice by honeybee swarms". In: *Philosophical transactions of The Royal Society B: biological sciences* 364.1518, pp. 755–762.

Lopez, Ugo et al. (2012). "From behavioural analyses to models of collective motion in fish schools". In: *Interface Focus* 2.6, pp. 693–707. ISSN: 2042-8898. DOI: 10.1098/rsfs.2012.0033. URL: https://royalsocietypublishing.org/doi/10.1098/rsfs.2012.0033.

Ma, Kevin Y et al. (2013). "Controlled flight of a biologically inspired, insect-scale robot". In: *Science* 340.6132, pp. 603–607.

Magurran, AE, WJ Oulton, and TJ Pitcher (1985). "Vigilant behaviour and shoal size in minnows". In: *Zeitschrift für Tierpsychologie* 67.1-4, pp. 167–178.

Martinez-Garcia, Ricardo, Cristóbal López, and Federico Vazquez (2021). "Species exclusion and coexistence in a noisy voter model with a competition-colonization tradeoff". In: *Physical Review E* 103.3, p. 032406.

Méndez, V., D. Campos, and F. Bartumeus (2014). *Stochastic Foundations in Movement Ecology*. Berlin, Heidelberg: Springer Verlag.

Mensi, Skander et al. (2012). "Parameter extraction and classification of three cortical neuron types reveals two distinct adaptation mechanisms". In: *Journal of neurophysiology* 107.6, pp. 1756–1775.

Mertens, Stephan and Cristopher Moore (2012). "Continuum percolation thresholds in two dimensions". In: *Physical Review E* 86.6, p. 061109.

Miguel, M Carmen et al. (2001). "Intermittent dislocation flow in viscoplastic deformation". In: *Nature* 410. DOI: 10.1038/35070524.

Miller, David L (2013). *Introduction to collective behavior and collective action*. Waveland Press.

Mishra, Shradha et al. (2012). "Collective dynamics of self-propelled particles with variable speed". In: *Physical Review E* 86.1, p. 011901.

Mitri, Sara et al. (2013). "Using robots to understand social behaviour". In: *Biological Reviews* 88.1, pp. 31–39.

Moiseff, Andrew and Jonathan Copeland (2010). "Firefly synchrony: a behavioral strategy to minimize visual clutter". In: *Science* 329.5988, pp. 181–181.

Molloy, M. and B. Reed (1995). "A critical point for random graphs with a given degree sequence". In: *Random Struct. Algorithms* 6, p. 161.

Mönck, Hauke Jürgen et al. (2018). "BioTracker: an open-source computer vision framework for visual animal tracking". In: *arXiv preprint arXiv:1803.07985*.

Mondada, Francesco et al. (2004). "SWARM-BOT: A new distributed robotic concept". In: *Autonomous robots* 17, pp. 193–221.

Mora, Thierry and William Bialek (2011). "Are biological systems poised at criticality?" In: *Journal of Statistical Physics* 144, pp. 268–302.

Moretti, Paolo and Miguel A Muñoz (2013). "Griffiths phases and the stretching of criticality in brain networks". In: *Nature communications* 4.1, p. 2521.

Múgica, Julia et al. (2022). "Scale-free behavioral cascades and effective leadership in schooling fish". In: *Scientific reports* 12.1, p. 10783.

Munoz, Miguel A (2018). "Colloquium: Criticality and dynamical scaling in living systems". In: *Reviews of Modern Physics* 90.3, p. 031001.

Murakami, Hisashi et al. (2015). "Inherent noise appears as a Lévy walk in fish schools". In: *Scientific Reports* 5.1, p. 10605. DOI: 10.1038/srep10605.

Nagy, Máté et al. (2010). "Hierarchical group dynamics in pigeon flocks." In: *Nature* 464.7290, pp. 890–893. ISSN: 0028-0836. DOI: 10.1038/nature08891. arXiv: 1010.5394.

Nagy, Máté et al. (2013). "Context-dependent hierarchies in pigeons." In: *Proceedings of the National Academy of Sciences of the United States of America* 110.32, pp. 13049–54. ISSN: 1091-6490. DOI: 10.1073/pnas.1305552110. URL: http://www.ncbi.nlm.nih.gov/pubmed/23878247.

Néda, Zoltán et al. (2000). "The sound of many hands clapping". In: *Nature* 403.6772, pp. 849–850.

Newman, Mark (2010). *Networks: An Introduction*. New York, NY, USA: Oxford University Press, Inc.

Pais, Darren et al. (2013). "A mechanism for value-sensitive decision-making". In: *PloS one* 8.9, e73216.

Palstra, Arjan P et al. (2010). "Establishing zebrafish as a novel exercise model: swimming economy, swimming-enhanced growth and muscle growth marker gene expression". In: *PLoS One* 5.12, e14483.

Panadeiro, Veronica et al. (2021). "A review of 28 free animal-tracking software applications: current features and limitations". In: *Lab animal* 50.9, pp. 246–254.

Parrish, Julia K and Leah Edelstein-Keshet (1999). "Complexity, pattern, and evolutionary trade-offs in animal aggregation". In: *Science* 284.5411, pp. 99–101.

Parrish, Julia K, Steven V Viscido, and Daniel Grunbaum (2002). "Self-organized fish schools: an examination of emergent properties". In: *The biological bulletin* 202.3, pp. 296–305.

Passino, Kevin M and Thomas D Seeley (2006). "Modeling and analysis of nest-site selection by honeybee swarms: the speed and accuracy trade-off". In: *Behavioral Ecology and Sociobiology* 59, pp. 427–442.

Patel, Aniruddh D et al. (2009). "Experimental evidence for synchronization to a musical beat in a nonhuman animal". In: *Current biology* 19.10, pp. 827–830.

Pearce, Daniel J. G. and Luca Giomi (2016). "Linear response to leadership, effective temperature, and decision making in flocks". In: *Phys. Rev. E* 94.2, p. 022612. ISSN: 2470-0045. DOI: 10.1103/PhysRevE.94.022612. URL: http://arxiv.org/abs/1511.03652http://link.aps.org/doi/10.1103/PhysRevE.94.022612.

Pérez-Escudero, Alfonso et al. (2014). "idTracker: tracking individuals in a group by automatic identification of unmarked animals". In: *Nature methods* 11.7, pp. 743–748.

Petit, Odile et al. (2009). "Collective decision-making in white-faced capuchin monkeys". In: *Proceedings of the Royal Society B: Biological Sciences* 276.1672, pp. 3495–3503.

Pfeifer, Rolf, Max Lungarella, and Fumiya Iida (2007). "Self-organization, embodiment, and biologically inspired robotics". In: *science* 318.5853, pp. 1088–1093.

Phan, Hoang Vu et al. (2019). "KUBeetle-S: An insect-like, tailless, hover-capable robot that can fly with a low-torque control mechanism". In: *International Journal of Micro Air Vehicles* 11, p. 1756829319861371.

Pitcher, TJ, AE Magurran, and IJ Winfield (1982). "Fish in larger shoals find food faster". In: *Behavioral Ecology and Sociobiology* 10.2, pp. 149–151.

Pitcher, Tony J (1983). "Heuristic definitions of fish shoaling behaviour." In: *Animal Behaviour*.

— (1998). "Shoaling and shoaling behaviour in fishes". In: *Comparative Psychology: A Handbook. New York: Garland*, pp. 748–760.

Pruessner, G. (2012). *Self-Organised Criticality: Theory, Models and Characterisation*. Self-organised Criticality: Theory, Models, and Characterisation. Cambridge, UK: Cambridge University Press. ISBN: 9780521853354. URL: https://books.google.es/books?id=TXKcpGqMSDYC.

Quera, Vicenç et al. (2019). "Local interaction rules and collective motion in black neon tetra (Hyphessobrycon herbertaxelrodi) and zebrafish (Danio rerio)." In: *Journal of Comparative Psychology* 133.2, p. 143.

Rands, Sean A et al. (2003). "Spontaneous emergence of leaders and followers in foraging pairs". In: *Nature* 423.6938, pp. 432–434.

Rao, Srinivasa R et al. (2019). "Small animal video tracking for activity and path analysis using a novel open-source multi-platform application (AnimApp)". In: *Scientific reports* 9.1, pp. 1–9.

Reina, Andreagiovanni et al. (2017). "Model of the best-of-N nest-site selection process in honeybees". In: *Physical Review E* 95.5, p. 052411.

Ribeiro, Haroldo V et al. (2022). "Effects of shady environments on fish collective behavior". In: *Scientific Reports* 12.1, p. 17873.

Rieucau, Guillaume, Lise Doksæter Sivle, and Nils Olav Handegard (2015). "Herring perform stronger collective evasive reactions when previously exposed to killer whales calls". In: *Behavioral Ecology* 27.2, pp. 538–544.

Rodriguez, Alvaro et al. (2018). "ToxTrac: a fast and robust software for tracking organisms". In: *Methods in Ecology and Evolution* 9.3, pp. 460–464.

Romanczuk, Pawel and Bryan C Daniels (2023). "Phase Transitions and Criticality in the Collective Behavior of Animals—Self-Organization and Biological Function". In: *Order, Disorder and Criticality: Advanced Problems of Phase Transition Theory*. World Scientific, pp. 179–208.

Romero-Ferrero, Francisco et al. (2019). "Idtracker. ai: tracking all individuals in small or large collectives of unmarked animals". In: *Nature methods* 16.2, pp. 179–182.

Rosenthal, Sara Brin et al. (2015). "Revealing the hidden networks of interaction in mobile animal groups allows prediction of complex behavioral contagion". In: *Proc. Natl. Acad. Sci.* 112.15, pp. 4690–4695. ISSN: 0027-8424. DOI: 10.1073/pnas.1420068112. URL: http://www.pnas.org/lookup/doi/10.1073/pnas.1420068112.

Rubenstein, Michael, Christian Ahler, and Radhika Nagpal (2012). "Kilobot: A low cost scalable robot system for collective behaviors". In: *2012 IEEE international conference on robotics and automation*. IEEE, pp. 3293–3298.

Rubenstein, Michael et al. (2013). "Collective transport of complex objects by simple robots: theory and experiments". In: *Proceedings of the 2013 international conference on Autonomous agents and multi-agent systems*, pp. 47–54.

Ruckstuhl, Kathreen E (1998). "Foraging behaviour and sexual segregation in bighorn sheep". In: *Animal behaviour* 56.1, pp. 99–106.

Ruxton, Graeme D (2012). "Group dynamics: predators and prey get a little help from their friends". In: *Current Biology* 22.13, R531–R533.

Sane, Sanjay P, Sree Subha Ramaswamy, and Sreekrishna Varma Raja (2020). "Insect architecture: structural diversity and behavioral principles". In: *Current Opinion in Insect Science* 42, pp. 39–46.

Sarfati, Raphaël, Julie C Hayes, and Orit Peleg (2021). "Self-organization in natural swarms of Photinus carolinus synchronous fireflies". In: *Science Advances* 7.28, eabg9259.

Sasaki, Takao and Stephen C Pratt (2018). "The psychology of superorganisms: collective decision making by insect societies". In: *Annual Review of Entomology* 63, pp. 259–275.

Savitzky, Abraham and Marcel JE Golay (1964). "Smoothing and differentiation of data by simplified least squares procedures." In: *Analytical chemistry* 36.8, pp. 1627–1639.

Schmickl, Thomas, Christoph Möslinger, and Karl Crailsheim (2007). "Collective perception in a robot swarm". In: *Swarm Robotics: Second International Workshop, SAB 2006, Rome, Italy, September 30-October 1, 2006, Revised Selected Papers 2*. Springer, pp. 144–157.

Seeley, Thomas D (1997). "Honey bee colonies are group-level adaptive units". In: *The American Naturalist* 150.S1, s22–S41.

— (2010). "Honeybee democracy". In: *Honeybee Democracy*. Princeton University Press.

Seeley, Thomas D and Susannah C Buhrman (2001). "Nest-site selection in honey bees: how well do swarms implement the" best-of-N" decision rule?" In: *Behavioral Ecology and Sociobiology* 49, pp. 416–427.

Seeley, Thomas D, Scott Camazine, and James Sneyd (1991). "Collective decision-making in honey bees: how colonies choose among nectar sources". In: *Behavioral Ecology and Sociobiology* 28, pp. 277–290.

Seeley, Thomas D and P Kirk Visscher (2004). "Quorum sensing during nest-site selection by honeybee swarms". In: *Behavioral Ecology and Sociobiology* 56, pp. 594–601.

Seeley, Thomas D et al. (2012). "Stop signals provide cross inhibition in collective decision-making by honeybee swarms". In: *Science* 335.6064, pp. 108–111.

Sharkey, Amanda JC (2006). "Robots, insects and swarm intelligence". In: *Artificial Intelligence Review* 26, pp. 255–268.

— (2007). "Swarm robotics and minimalism". In: *Connection Science* 19.3, pp. 245–260.

Shen, Jackie (Jianhong) (2008). "Cucker–Smale Flocking under Hierarchical Leadership". In: *SIAM Journal on Applied Mathematics* 68.3, pp. 694–719. DOI: 10.1137/060673254. eprint: https://doi.org/10.1137/060673254. URL: https://doi.org/10.1137/060673254.

Sih, Andrew, Sean F Hanser, and Katherine A McHugh (2009). "Social network theory: new insights and issues for behavioral ecologists". In: *Behavioral Ecology and Sociobiology* 63.7, pp. 975–988.

Silverberg, Jesse L et al. (2013). "Collective motion of humans in mosh and circle pits at heavy metal concerts". In: *Physical review letters* 110.22, p. 228701.

Slavkov, Ivica et al. (2018). "Morphogenesis in robot swarms". In: *Science Robotics* 3.25, eaau9178.

Smith, Jennifer E. et al. (2016). "Leadership in Mammalian Societies: Emergence, Distribution, Power, and Payoff". In: *Trends in Ecology & Evolution* 31.1, pp. 54–66. ISSN: 01695347. DOI: 10.1016/j.tree.2015.09.013. URL: http://dx.doi.org/10.1016/j.tree.2015.09.013https://linkinghub.elsevier.com/retrieve/pii/S0169534715002499.

Sobral, Andrews and Antoine Vacavant (2014). "A comprehensive review of background subtraction algorithms evaluated with synthetic and real videos". In: *Computer Vision and Image Understanding* 122, pp. 4–21.

Sood, Vishal and Sidney Redner (2005). "Voter model on heterogeneous graphs". In: *Physical review letters* 94.17, p. 178701.

Sornette, Didier (2006). *Critical phenomena in natural sciences: chaos, fractals, selforganization and disorder: concepts and tools*. Springer Science & Business Media.

Sosna, Matthew MG et al. (2019). "Individual and collective encoding of risk in animal groups". In: *Proceedings of the National Academy of Sciences* 116.41, pp. 20556–20561.

Spivak, Marla and Robert G Danka (2021). "Perspectives on hygienic behavior in Apis mellifera and other social insects". In: *Apidologie* 52, pp. 1–16.

Sridhar, Vivek Hari, Dominique G Roche, and Simon Gingins (2019). "Tracktor: image-based automated tracking of animal movement and behaviour". In: *Methods in Ecology and Evolution* 10.6, pp. 815–820.

Stauffer, Dietrich and Ammon Aharony (2018). *Introduction to percolation theory*. Taylor & Francis.

Stork, Nigel E (2018). "How many species of insects and other terrestrial arthropods are there on Earth?" In: *Annual review of entomology* 63, pp. 31–45.

Strandburg-Peshkin, Ariana et al. (2018). "Inferring influence and leadership in moving animal groups". In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 373.1746, p. 20170006. ISSN: 0962-8436. DOI: 10.1098/rstb.2017.0006.

Sumpter, D. J. (2010). *Collective Animal Behavior*. New Jersey: Princeton University Press.

Sumpter, D. J. T. (2006). "The principles of collective animal behaviour". In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 361.1465, pp. 5–22. ISSN: 0962-8436. DOI: 10.1098/rstb.2005.1733. URL: http://rstb.royalsocietypublishing.org/cgi/doi/10.1098/rstb.2005.1733.

Sumpter, David JT et al. (2008). "Consensus decision making by fish". In: *Current Biology* 18.22, pp. 1773–1777.

Teixeira, TF et al. (2016). "Two new species of Hyphessobrycon (Characiformes: Characidae) from the headwaters of the Tapajós and Xingu River basins, Pará, Brazil". In: *Journal of Fish Biology* 88.2, pp. 459–476.

Tharin, Julien (n.d.). *Kilobot User Manual*. English. Version 2.0. K-Team. 44 pp.

Thrun, Michael C and Alfred Ultsch (2021). "Swarm intelligence for self-organized clustering". In: *Artificial Intelligence* 290, p. 103237.

Trianni, Vito et al. (2016). "Emergence of consensus in a multi-robot network: from abstract models to empirical validation". In: *IEEE Robotics and Automation Letters* 1.1, pp. 348–353.

Tunstrøm, Kolbjørn et al. (2013). "Collective states, multistability and transitional behavior in schooling fish". In: *PLoS computational biology* 9.2, e1002915.

Valentini, Gabriele, Eliseo Ferrante, and Marco Dorigo (2017). "The best-of-n problem in robot swarms: Formalization, state of the art, and novel perspectives". In: *Frontiers in Robotics and AI* 4, p. 9.

Valentini, Gabriele, Heiko Hamann, Marco Dorigo, et al. (2014). "Self-organized collective decision making: the weighted voter model." In: *AAMAS*, pp. 45–52.

Valentini, Gabriele et al. (2016). "Collective decision with 100 Kilobots: Speed versus accuracy in binary discrimination problems". In: *Autonomous agents and multi-agent systems* 30.3, pp. 553–580.

Varela, Francisco J and Paul Bourgine (1992). *Toward a practice of autonomous systems: Proceedings of the First European Conference on Artificial Life*. MIT press.

Vicsek, Tamás and Anna Zafeiris (2012). "Collective motion". In: *Phys. Rep.* 517.3-4, pp. 71–140. ISSN: 03701573. DOI: 10.1016/j.physrep.2012.03.004. URL: http://linkinghub.elsevier.com/retrieve/pii/S0370157312000968.

Vicsek, Tamas et al. (1995). "Novel Type of Phase Transition in a System of Self-Driven Particles". In: *Phys. Rev. Lett.* 75, pp. 1226–1229. DOI: 10.1103/PhysRevLett.75.1226.

Videler, JJ and CS Wardle (1991). "Fish swimming stride by stride: speed limits and endurance". In: *Reviews in Fish Biology and Fisheries* 1, pp. 23–40.

Videler, John J and D Weihs (1982). "Energetic advantages of burst-and-coast swimming of fish at high speeds". In: *Journal of Experimental Biology* 97.1, pp. 169–178.

Viscido, Steven V, Julia K Parrish, and Daniel Grünbaum (2004). "Individual behavior and emergent properties of fish schools: a comparison of observation and theory". In: *Marine Ecology Progress Series* 273, pp. 239–249.

Walt, Stefan Van der et al. (2014). "scikit-image: image processing in Python". In: *PeerJ* 2, e453.

Walter, Tristan and Iain D Couzin (2021). "TRex, a fast multi-animal tracking system with markerless identification, and 2D estimation of posture and visual fields". In: *Elife* 10, e64000.

Wang, Xin-Guang et al. (2013). "A modified Vicsek model for self-propelled agents with exponential neighbor weight and restricted visual field". In: *Physica A: Statistical Mechanics and its Applications* 392.10, pp. 2398–2405.

Ward, Ashley J. W. et al. (2011). "Fast and accurate decisions through collective vigilance in fish shoals". In: *Proceedings of the National Academy of Sciences* 108.6, pp. 2312–2315. DOI: 10.1073/pnas.1007102108. URL: http://www.pnas.org/content/108/6/2312.abstract.

Webb, Barbara (2001). "Can robots make good models of biological behaviour?" In: *Behavioral and brain sciences* 24.6, pp. 1033–1050.

Webb, P. W. (1984). "Body Form, Locomotion and Foraging in Aquatic Vertebrates". In: *American Zoologist* 24.1, pp. 107–120. ISSN: 00031569. URL: http://www.jstor.org/stable/3882756 (visited on 05/23/2022).

Weidenmüller, Anja (2004). "The control of nest climate in bumblebee (Bombus terrestris) colonies: interindividual variability and self reinforcement in fanning response". In: *Behavioral Ecology* 15.1, pp. 120–128.

Weihs, D (1973). "Hydromechanics of fish schooling". In: *Nature* 241.5387, pp. 290–291.

Welch, P. (1967). "The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms". In: *IEEE Transactions on Audio and Electroacoustics* 15.2, pp. 70–73.

Werkhoven, Zach et al. (2019). "MARGO (Massively Automated Real-time GUI for Object-tracking), a platform for high-throughput ethology". In: *PloS one* 14.11, e0224243.

Wild, Benjamin et al. (2021). "Social networks predict the life and death of honey bees". In: *Nature communications* 12.1, pp. 1–12.

Wilson, Edward Osborne (1971). "Social insects". In: *Science* 172.3981, pp. 406–406.

Wischmann, Steffen et al. (2006). "Synchronization of internal neural rhythms in multi-robotic systems". In: *Adaptive Behavior* 14.2, pp. 117–127.

Wu, Guanhao, Yan Yang, and Lijiang Zeng (June 2007). "Kinematics, hydrodynamics and energetic advantages of burst-and-coast swimming of koi carps (Cyprinus carpio koi)". In: *Journal of Experimental Biology* 210.12, pp. 2181–2191. ISSN: 0022-0949. DOI: 10.1242/jeb.001842. eprint: https://journals.biologists.com/jeb/article-pdf/210/12/2181/1259194/2181.pdf. URL: https://doi.org/10.1242/jeb.001842.

Xu, Zhiping and Xi En Cheng (2017). "Zebrafish tracking using convolutional neural networks". In: *Scientific reports* 7.1, pp. 1–11.

Yang, Xiufeng, Longlong Chang, and Néstor O Pérez-Arancibia (2020). "An 88-milligram insect-scale autonomous crawling robot driven by a catalytic artificial muscle". In: *Science Robotics* 5.45, eaba0015.

Yeomans, J. M. (1992). *Statistical mechanics of phase transitions*. Oxford: Oxford University Press.

Zapperi, Stefano et al. (1998). "Dynamics of a ferromagnetic domain wall: Avalanches, depinning transition, and the Barkhausen effect". In: *Phys. Rev. B* 58 (10), pp. 6353–6366. DOI: 10.1103/PhysRevB.58.6353. URL: https://link.aps.org/doi/10.1103/PhysRevB.58.6353.

Zapperi, Stefano et al. (1999). "Avalanches in breakdown and fracture processes". In: *Phys. Rev. E* 59 (5), pp. 5049–5057. DOI: 10.1103/PhysRevE.59.5049. URL: https://link.aps.org/doi/10.1103/PhysRevE.59.5049.