The role of parasitism in individual and collective behavior of guppies

Doctoral thesis for obtaining the academic degree

Doctor of Natural Sciences
Dr.rer.nat.

submitted by

Albi, Angela

at the

Universität Konstanz

Faculty of Sciences

Department of Biology

Konstanz, 2024

Date of examination: February 7, 2023
1. Reviewer: Iain Couzin
2. Reviewer: Sandra Binning

Contents

Summary

Collective behavior is a fascinating and ubiquitous phenomenon in the animal kingdom, with many examples, including the coordinated motion of cells, schooling fish, and flocking birds. Among organisms, collective action can benefit group members, such as by enhancing their ability to avoid predators or find food. In addition, social interactions can be influenced by other factors, such as changes in temperature or resources, or the presence of pathogens. The study of collective behavior can help us understand how diseases spread through populations, and by establishing how individual animals respond to a pathogen, we might be able to better predict how a disease can spread.

With recent technological advances in image-based analysis software and algorithms, we can study animal behavior in greater detail than ever before, and the use of such quantitative methodology allows researchers to uncover new insights into how and why animals behave in the way they do. In this work, we employed cutting-edge tracking software technology to study how the behavior of guppies (*Poecilia reticulata*) is affected by the infection of the socially transmitted ectoparasite *Gyrodactylus* sp. both at an individual and at a collective level. In Chapter 1, we use a swim tunnel with intermittent flow respirometry and develop a quantitative approach that allows us to combine information regarding swimming kinematics, performance, and respirometry to evaluate the effect of parasitism on individual swimming behavior across increasing water flow speeds. Guppies are highly social; therefore, in Chapter 2 we extend our questions to the collective level, presenting experiments designed to investigate how the presence of a contagious parasite affects the fission-fusion dynamics of freely shoaling guppies move and interact. Finally, in Chapter 3, we investigate how differential social influence (also referred to as leader-follower dynamics) varies across a broad range of interaction timescales of interactions, and evaluate how this relates to parasitism.

The combined results of these analyses demonstrate the value of an integrative approach to better understand the role of disease across scales from individual to group - and from group to inter-group - dynamics in the study of animal behavior.

Zusammenfassung

Mit den jüngsten technologischen Fortschritten bei bildbasierter Analysesoftware und Algorithmen können wir das Verhalten von Tieren detaillierter als je zuvor untersuchen. In dieser Arbeit haben wir modernste Tracking-Softwaretechnologie eingesetzt, um zu untersuchen, wie das Verhalten von Guppys (*Poecilia reticulata*) durch die Infektion mit dem sozial übertragenen Ektoparasiten *Gyrodactylus* sp. beeinflusst wird, sowohl auf individueller als auch auf kollektiver Ebene. In Kapitel 1 verwenden wir einen Schwimmtunnel mit intermittierender Strömungsatmung und entwickeln einen quantitativen Ansatz, der es uns ermöglicht, Informationen über Schwimmkinematik, Leistung und Atmung zu kombinieren. Auf der Grundlage dieser Informationen untersuchen wir die Auswirkungen von Parasitismus auf das individuelle Schwimmverhalten bei steigenden Strömungsgeschwindigkeiten. Guppys sind hochgradig soziale Tiere; daher dehnen wir in Kapitel 2 unsere Fragestellungen auf ihr kollektives Verhalten aus und stellen Experimente vor, mit denen wir untersuchen wollen, wie sich die Anwesenheit eines ansteckenden Parasiten auf die Spaltungs- und Fusionsdynamik von frei schwimmenden Guppys auswirkt und wie sie sich bewegen und interagieren. In Kapitel 3 untersuchen wir schließlich, wie sich der unterschiedliche soziale Einfluss (auch als Anführer-Nachfolger-Dynamik bezeichnet) über ein breites Spektrum von Interaktionszeitskalen hinweg verändert, und bewerten, wie dies mit Parasitismus zusammenhängt.

Die kombinierten Ergebnisse dieser Analysen zeigen den Stellenwert eines integrativen Ansatzes bei der Untersuchung von Tierverhalten für ein besseres Verständnis der Rolle von Krankheiten über Skalen hinweg - vom Individuum bis zur Gruppe und von der Gruppe bis zur Intergruppendynamik.

General Introduction

The movement and behavior of animals in groups have attracted researcher interest for many decades. This work has focused on understanding how groups function and what mechanisms drive emergent collective behaviors, as coordinated group patterns that result from individual interactions (Couzin, 2007; Strandburg-Peshkin et al., 2015; Sumpter, 2010). From ants to fish to primates, there exist remarkable examples of emergent properties that shape how animals make decisions, forage, escape predators, communicate, mate, move, and interact with the surrounding environment. For example, by swarming together, bees use a form of collective intelligence to effectively find a new home (Seeley, 2010). Similarly, fish school together to improve their chances of survival by confusing predators and making it harder for them to single out an individual fish (Ioannou et al., 2015; Sumpter, 2006).

Unfortunately, group living is not without costs. As a consequence of the COVID-19 pandemic, there is renewed interest in understanding how groups of highly connected individuals respond to the threat of disease from parasites or other contagious pathogens, and the implications of collective action on disease spread (Lopes, 2022; Lopes et al., 2021; Stockmaier et al., 2021). Although often overlooked, parasites are ubiquitous and highly successful disease-causing organisms (Dobson et al., 2008; Poulin and Morand, 2000), and can impact the biology of the host at different levels, playing an important role in the life of an animal (Marcogliese, 2004,0; Timi and Poulin, 2020).

Parasitism is a symbiotic relationship where one organism, the parasite, lives in or on another organism, the host, causes some degree of damage and lives at the expense of the host (Combes, 2001; Price, 1977; Raffel et al., 2008). Viruses, bacteria, fungi, protozoa, as well as worm and arthropod endo- or ecto- macro-parasites can all act as parasites at some stage in their lives. The diversity of parasites also mirrors the variety of defensive strategies that both vertebrates and invertebrates adopt to reduce infection risk or damage (Curtis, 2014). For example, the response strategies of both infected and uninfected hosts can include tolerance - when the host limits and ameliorates the damage of the infection, or resistance - when the host actively fights against, or avoidance - when a host (or group) takes action to reduce the likelihood of an infection (Best et al., 2008;

Curtis, 2014; Stephenson and Adelman, 2022). However, avoidance of parasites, like the avoidance of predators, can have associated costs (Buck et al., 2018). For instance, animals may avoid close contact with conspecifics that show signs of infection and such avoidance could increase predation risk, introducing a clear trade-off (Buck et al., 2018). The choice of a solitary lifestyle is not risk-free (Bordes et al., 2007), and therefore other avoidance strategies are used by group-living animals. In addition to the often cited avoidance mechanisms, self-sorting mechanisms are also likely involved in changes in individual to group movement patterns or behavioral motivations, due to parasitic infections. Differences among individuals (*e.g.* behavioral state, age, size) can influence the position occupied within a group, which results in self-sorting (Couzin and Krause, 2003; Couzin et al., 2002). Therefore, if physical impairment due to infection can change host speed, acceleration, and turning, it might, in turn, alter its spatial positioning within a group thus affecting overall group dynamics (Couzin and Krause, 2003; Couzin et al., 2002).

One mechanism of parasite avoidance is detection by the uninfected host (Barber et al., 2000; Behringer et al., 2018). Detection is achieved via sensory cues that are generally divided between visual, olfactory, mechanosensory, chemosensory, tactile and acoustic, or a combination of these (Barber et al., 2000; Behringer et al., 2006; Binning et al., 2013; Dugatkin et al., 1994; Jolles et al., 2020b; Kavaliers and Choleris, 2018; Kavaliers et al., 2004; Krause, 1993; Lopes, 2022; Lopes et al., 2022; McElroy and de Buron, 2014; Rahn et al., 2015; Stephenson, 2019; Stephenson et al., 2018; Sumpter et al., 2008; Timi and Poulin, 2020; Ward et al., 2002; Östlund Nilsson et al., 2005). Each of the cues that could play a role in detection depends on the environment, as well as the parasite life cycle. For example, rodents infected with endoparasites have been previously shown to rely heavily on olfactory cues to avoid parasitized conspecifics (Kavaliers and Choleris, 2018; Kavaliers et al., 2004), primates and some fishes seem to use a combination of visual and olfactory cues (Dugatkin et al., 1994; Stephenson, 2019; Stephenson et al., 2018), while birds seem to rely mainly on visual cues (Spurrier et al., 1991).

Another often unexplored mechanism for the detection of parasitism is that of group facilitation. Group-living animals can pool information of group members to reach more accurate decisions (King and Cowlishaw, 2007; Simons, 2004), and it was previously found that sticklebacks make more accurate decisions when choosing whether to move towards healthy versus sick-looking fish when part of a group of eight individuals, compared with smaller group sizes, using a quorum consensus, which was found to enable individuals to better avoid apparently-diseased individuals (Sumpter et al., 2008).

In the context of social behavior and parasitic infection, in this thesis, we examine how individual swimming performance and the collective behavior of a social fish

species, the guppy (*Poecilia reticulata*), are affected by infection with the monogenean ectoparasite *Gyrodactylus* sp. Among Trinidadian guppies, gyrodactylids are one of the most prevalent parasitic worms (Mohammed et al., 2020; Stephenson et al., 2015). *Gyrodactylus* sp. was previously shown to affect feeding rate (Van Oosterhout et al., 2003), males' coloration (Houde and Torio, 1992), survival and reproductive fitness (van Oosterhout et al., 2007). Also, heavy infections can damage the epidermis, which could affect drag and thus alter swimming performance. It has also been previously suggested that parasite infection might affect swimming mechanics, making infected fish swim 'erratically' (Bakke et al., 2007; Cable et al., 2002), which could be an effect of high loads of the parasites causing damage or changes in the shape of the fins of the host (Cable et al., 2002). However, to date, there has not been a quantification of the effects of *Gyrodactylus* sp. infection on the kinematics of host swimming behavior.

A few studies have previously quantified healthy guppy's swimming behavior and locomotion patterns (Killen et al., 2016; Svendsen et al., 2013), and one study has looked at the effect of *Gyrodactylus* sp. infection on turbulent flow swimming and how this may affect parasite transmission rates (Reynolds et al., 2019). Locomotion is an important trait for guppies as they live in pools and streams with variable flow regimes. The ability to maintain a steady and stable swimming position in turbulent water is an essential survival strategy to both avoid being washed downstream and to control position relative to other individuals (Hockley et al., 2014). Previous studies on guppies have looked at the link between locomotion and other factors such as reproductive allocation, male coercive behaviors, and male coloration (Banet et al., 2016; Killen et al., 2016; Nicoletto, 1991; Svendsen et al., 2013). In particular, increased coercive behavior by male guppies can lead to increased swimming efficiency in female guppies, also associated with a decrease in the pectoral fin use (Killen et al., 2016). Similarly, excessive pectoral fin use can result in increased swimming costs and lower maximum critical speed (U_{crit}) (Killen et al., 2016; Svendsen et al., 2013). However, it remains unclear whether the ectoparasite *Gyrodactylus* sp. has an effect on individuals' swimming abilities and locomotion, and whether altered swimming behavior possibly contributes to the mechanism of avoidance by visual cues.

In Chapter 1 we quantify how locomotion, oxygen consumption, and swimming performance of individual guppies vary in response to parasitic infection. The goal of this study is to understand whether *Gyrodactylus* sp. infection causes changes in individual swimming behavior and kinematics, which could inform us about possible cues involved in visual-based avoidance mechanisms previously reported in the same system (Stephenson, 2019). We present the results of experiments performed on 46 guppies, divided into infected and uninfected treatments. We find that by quantifying locomotion,

oxygen consumption rate, and swimming performance traits measured at increasing water flows, we can discriminate between fish carrying different parasite infection loads. However, fish carrying different parasite loads (as well as unparasitized fish) show some overlap in their traits, which reflects a more general concept of swimming kinematics as a continuum rather than discrete categories of swimming modes whether infected or not (Di Santo et al., 2021).

Guppies are a social species and live in groups. Therefore, in Chapter 2, we use the same host-parasite system to understand how parasites affect the movement dynamics of infected individuals, and how groups of guppies respond to the presence of a parasitized individual in the shoal. Previous studies on the guppy-*Gyrodactylus* system in the context of social behavior showed that guppies avoid infected conspecifics using a combination of olfactory and visual cues (Stephenson, 2019), and that predation pressure in the guppy natural habitat plays an important role in selecting on parasite virulence (Walsman et al., 2022). Moreover, *Gyrodactylus* sp. infection was shown to modify fission-fusion dynamics, by increasing the number of initiated shoal fissions (group splitting into smaller groups), possibly to reduce the likelihood of being in contact with the infected hosts (Croft et al., 2011). However, these experiments failed to obtain the information on the two-way dynamical interactions that we achieve by recording detailed individual- and group-level information of each individual in the group. We designed an experiment to record eight guppies freely shoaling together for three hours, manipulating the infection status of one of the shoal members. Using cutting-edge tracking technology (Walter and Couzin, 2021), we maintain the individual identity of each unmarked fish throughout the entire experiment duration. This approach allows us to quantify in detail the interactions among shoal members, as well as quantify individual motion features (such as speed, acceleration, and turning rates) that can vary with parasitic infection and possibly have cascading effects on group social behavior. This design overcomes the limitations of a binary-choice experimental design (previously used for example in Stephenson (2019)) and complements previous observational approaches Croft et al. (2011) (where fish were observed for a period of 30 minutes).

The results of Chapter 2 show that the effect of parasite infection is most clearly seen in individual behavioral patterns, rather than the overall group-level behaviors. This suggests that an individual-based approach is essential to investigate the effects of parasitic infection on group behavior. More specifically, we find that infected guppies swim at higher nearest-neighbor distances than uninfected conspecifics, spend more time in isolation, and are more likely to occupy more peripheral and frontal positions within the group. Overall, with this chapter, we discuss the possible mechanisms that underlie behavioral modifications and compare this to the other findings in the literature

on parasitism in the social context (Croft et al., 2011; Hoare et al., 1998; Krause et al., 1992; Stephenson, 2019; Ward et al., 2002).

Finally, in the third Chapter, we explore a relatively understudied aspect of collective behavior: the dynamics of inter-individual leader-follower relationships across considerably different timescales. Using this approach, we aim to better understand the role of individual variation in the group (*e.g.* in the context of parasitism), and how this contributes to group formation patterns, for example between highly cohesive or highly uncorrelated individuals of a shoal.

Previous studies have investigated the structure of animal groups, revealing patterns in the organization that reflect either hierarchical organizations, as opposed to egalitarian organizations (Dyer et al., 2009; Nagy et al., 2010). In the first case, when groups are hierarchical in their coordination of motion, individuals can take roles that are maintained over the course of an interaction. For example, in pigeon flocks, individual directional choices are governed by clear leader-follower relationships. This means that by properties of motions of short-duration interactions (*e.g.* velocity), some animals lead the direction of the movement while others reliably adjust their motion and copy it (Nagy et al., 2010; Sridhar, 2022). Here, by contrast to previous studies that mainly analyzed near-instantaneous temporal relationships, we expand the analysis to also describe leader-follower dynamics that range from fast-timescale speed adjustments to those that occur over minutes. We find consistent inter-group differences and we describe differences in leader-follower relationships of the tested groups at the short versus long timescale. Finally, we discuss preliminary data on how pairwise leader-follower dynamics across timescales also vary as a function of parasite load. We find that the magnitude of leader-follower scores (how leader or follower an individual is) decreases for guppies with high parasite loads, together with an overall decrease of the correlation of speed and activity levels with the other shoal members.

To conclude, our work provides new insights regarding the effect that *Gyrodactylus* sp. infection has on both individual and collective behavior of guppies, from physiology and locomotion to fission-fusion dynamics to leader-follower relationships across temporal scales. Parasitism is a pervasive phenomenon and the approach we developed in Chapter 1 can be applied to quantify locomotion and respirometry across many fish species, to characterize the swimming phenotypic variation that is not restricted to the effect of parasitism. The method developed for Chapters 2 and 3 can be applied to other fission-fusion systems or collectives, with an approach that highlights both the role of the individual and the dynamic social context; this approach can aid parasite studies as well as studies that investigate how the internal state of one or more individuals, affect the group coordination and movement. We think we can generalize our findings on in-

dividual variation and group motion to similar host-parasite systems, in particular for fission-fusion animal groups in the presence of a contagious parasite; however, this is yet to be tested. Overall, the study of parasitism is important not only for understanding the ecology of individual species but also for understanding the evolution of cooperation and conflict between different species. By understanding how and why parasites affect their hosts, we can learn more about the intricate web of interactions that make up the natural world.

Chapter 1

The effect of parasitism on locomotion and physiology in guppies: a quantitative approach

1.1 Introduction

The majority of animals in natural populations are infected with parasites (Curtis, 2014; Timi and Poulin, 2020), and an individual's response to parasite exposure and infection is often multidimensional, linked to phenotypic changes, which can have cascading impacts at ecological and evolutionary levels (Barber et al., 2000; Binning et al., 2013; Jolles et al., 2020b; McElroy and de Buron, 2014). Parasites live in or on their hosts, and can affect physiology, morphology, reproduction and behavior (Timi and Poulin, 2020). Because infections often negatively impact the host, many animals use sensory cues to detect the presence of pathogens and limit their exposure. For instance, parasites can affect movement, vocalization, or smell (Barber et al., 2000; Binning et al., 2013; Jolles et al., 2020b; Kavaliers and Choleris, 2018; Kavaliers et al., 2004; Lopes et al., 2022; McElroy and de Buron, 2014; Timi and Poulin, 2020; Östlund Nilsson et al., 2005), which, in principle, could be used by conspecifics to inform a behavioral response. For example, the bullfrog tadpoles *Rana catesbeiana* use chemical cues to detect conspecifics infected by the pathogenic yeast *Candida humicola* (Kiesecker et al., 1999). Using visual cues, female guppies show a preference for males with fewer parasites (nematode *Camallanus cotti* or monogenean *Gyrodactylus* sp.), as infected males show lower rates of mating display (Kennedy et al., 1987). Although the consequences of parasites can highly impact the behavior of the infected hosts and have consequences on group living, only a few studies have examined the effects of parasites on locomotion patterns.

Fish use different forms of locomotion at different swimming speeds (Lauder, 2015),

a strategy that might have evolved to minimize power requirements at a given swim speed, thus optimizing the cost of transport over a range of speeds (Korsmeyer et al., 2002). Water flows generate body displacement, expressed as a linear or angular translation of the body, and can have different effects on fish behavior. Kinematics refers to any change in body motion relative to the fish frame of reference, and evaluating the effect of parasite infection on kinematics and performance may inform us regarding aspects impacting both host and parasite fitness (Umberger et al., 2013). For example, differences in movement patterns could provide a reliable cue of infection, and thus drive avoidance behavior by conspecifics. In aquatic environments, visual cues have been previously shown to play a major role in parasite avoidance in killifish, sticklebacks and guppies (Dugatkin et al., 1994; Krause, 1993; Rahn et al., 2015; Stephenson, 2019; Stephenson et al., 2018; Sumpter et al., 2008; Ward et al., 2002) where infected fish have been reported to swim 'erratically', a visually-conspicuous behavior, which suggests swimming locomotion may play an important role in the recognition of infected individuals by conspecifics and thus modify shoaling decisions (Bakke et al., 2007; Barber et al., 2017; Cable et al., 2002; Dugatkin et al., 1994; Hockley et al., 2014; Krause and Godin, 1996). Similarly, guppies infected with gyrodactylids also show abnormal behaviors, such as "rubbing up" against shoal mates (Croft et al., 2011); moreover, in later stages of infection, host fins can become contracted and fin rays fuse together, which could indirectly modify their swimming patterns and efficiency (Cable et al., 2002; Hockley et al., 2014). Besides avoidance by healthy conspecifics, changes in locomotion patterns might also affect an individual's position relative to conspecifics via self-sorting mechanisms (Couzin and Krause, 2003). For example, a decrease in individual speeds and performance of infected individuals may account for differences in group shape and schooling dynamics (Couzin and Krause, 2003; Couzin et al., 2002; Jolles et al., 2017; Umberger et al., 2013; Ward et al., 2002). Therefore, understanding how parasites affect the swimming behavior and locomotion of the host is essential to understanding the phenotypic correlation between infection, performance, and social behavior.

Fish locomotion is often studied in the context of various biotic and abiotic factors, including predation, morphology, mating, water flow, foraging, or urbanization (Di Santo et al., 2017; Domenici et al., 2008; Kern and Langerhans, 2019; Killen et al., 2016; Liao, 2007). Changes in water flow regime are known to cause direct and indirect effects on fish behavior, physiology, and swimming ability (Binning et al., 2014,0; Langerhans, 2008; Liao, 2007; Nadler et al., 2018; Reynolds et al., 2019; Roche et al., 2013b) as well as on shoaling dynamics (Ashraf et al., 2017; Li et al., 2020; Marras et al., 2015), and can influence the selection of traits within the context of the environment (Langerhans, 2008). The way in which water flows affects behavior has also been studied in relationship

to host-parasite interactions. For example, Binning et al. (2013) showed that ectoparasites increase swimming costs of the coral reef fish *Scolopsis bilineata* tested at different water speeds, with infected fish showing increased standard metabolic rates, lower aerobic scope, and lower maximum swimming speeds than uninfected fish (Binning et al., 2013). Since they attach externally to their hosts, ectoparasites can increase drag and thus locomotor costs. With a similar approach, in this study, we compare the swimming behavior of parasite-uninfected and parasite-infected guppies, through measures of physiology (oxygen consumption rates), kinematics, and performance (maximum critical speed) across increasing flow speeds. Understanding and predicting trait changes due to infection in response to water flow can increase our understanding of how parasite infection interacts with physiology and locomotion, which can have direct effects on mechanisms of shoaling, transmission, avoidance, and self-sorting.

In this study, we compare swimming respirometry, locomotion, and performance across different flow regimes of non-parasitized (N=20) and parasitized fish with different infection loads (N=26). All locomotion parameters are estimated in increasing water flow speeds. Flow speed is controlled and oxygen consumption rate (MO2) is measured while the fish is swimming against the current in laminar flow. As a second estimate of oxygen use, we use the Factorial Aerobic Scope (FAS), a measure of the proportional increase of oxygen use across flow speeds (Clark et al., 2013). By measuring the oscillations properties of the body wave, we calculate Head:Tail (H:T) amplitude ratio, Strouhal number, tail curvature, length of the propulsive body wave, and pectoral fin use (Di Santo et al., 2021; Svendsen et al., 2013). Altogether, these measures describe variations in fish's undulatory propulsion, swimming efficiency, and biomechanics. As further indicators of swimming ability, we distinguish between rates of steady and unsteady swimming, and the relative position in the tank (front-back) (Langerhans, 2009; Roche et al., 2013b). Steady swimming (cruising) is characterized by periodic oscillations and constant speed, while unsteady swimming includes transient speed and unstable locomotor patterns, including rapid turns, fast starts, burst-and-coast swimming (Langerhans, 2009). Morphology was quantified using two metrics, the Scaled Mass Index (SMI) and the Fineness ratio, which are measures to estimate body condition (SMI) and shape (Fineness ratio) (Maceda-Veiga et al., 2014; Peig and Green, 2009). Finally, we estimated swimming performance by measuring the maximum swimming critical speed U_{crit}, which is the swimming speed at which fish stop swimming by exhaustion (Roche et al., 2013a).

Using a supervised dimensionality reduction method, Linear Discriminant Analysis (LDA) we discriminate individuals carrying different infection loads and we describe the features that contribute to distinguishing the groups. Moreover, the different classes

show a level of overlap that reflects a more general concept of swimming kinematics as a continuum rather than discrete categories (Di Santo et al., 2021), which highlights the importance of a more holistic evaluation of swimming performance and its multi-dimensional traits to better understand how fish can modify some kinematic variables to modulate their motion. Besides distinction among parasite load classes, we generalize our results using Principal Component Analysis (PCA) and clustering by describing swimming modes that reflect differences in swimming kinematics, morphology, and performance. This suggests that even in the absence of discrete swimming mode categories, by using a combination of metrics we can find within-species variation that informs us regarding phenotypic inter-individual variation that can be linked to possible different behavioral types. Overall, the results show that the effect of parasitism on locomotion is not uni-directional and the quantitative approach we developed describes the range of variation of swimming behavior in relation to parasitism, locomotion, morphology, and physiology.

1.2 Methods

1.2.1 Study system

We used Trinidadian Guppies *Poecilia reticulata*, held in the facilities of the Max Planck Institute of Animal Behavior, Konstanz, Germany. The original stock was received in 2017 by Manfred Schartl from the Univerisity of Würzburg. Holding tanks had circulating water, fish were maintained at 26.5±1 ℃, with average 7.5 pH and 400 μS/cm conductance, under a 12:12 light:darkness cycle and fed daily on live high HUFA *Artemia*, frozen food, including frozen cyclops, frozen *Artemia* and dry food (Naturefood, Supreme Plant - XS and S). All experiments were done in accordance with the permit AZ 35-9185.81/G-20/11 granted by the Regierungspräsidium Freiburg, Baden-Württembeg, Germany.

1.2.2 Infection and screening

For the experimental infections, we initiated a parasite culture of *Gyrodactylus* sp. which were first collected from a local pet shop, and subsequently maintained on guppies from the same experimental fish stock (culture). After each experimental infection, fish were held in a quarantine room, housed individually at a density of 13L per fish in 80L tanks divided in 6 compartments. The compartments were separated by transparent partitions to allow the fish to see conspecifics (thus called semi-isolation). Partitions had holes to allow water to flow across compartments. Holding conditions were identical for both experimentally infected fish and control sham-infected fish. Control and Treatment tri-

als were conducted at different moments to avoid holding unparasitized and parasitized fish in the same system.

Wild caught Trinidadian guppies generally host loads of about 10 worms/host of Gyrodactylus sp. (Harris and Lyles, 1992), but from field surveys, guppies were found to host up to about 270 parasites (Stephenson et al., 2015). *Gyrodactylus* sp. is a monogenean flatworm that reproduces on the skin of the host with a generation time of 24h. Parasite load grows over time but the load on hosts can vary across individuals (Stephenson, 2019; Stephenson et al., 2017), for example with respect to body size, with larger individuals hosting higher parasite loads (Cable and van Oosterhout, 2007; Richards et al., 2010; van Oosterhout et al., 2008).

To initiate infection, first, a culture fish with a heavy infection of worms on the caudal fin was euthanized with an overdose of 0.25 mg/ml MS222 (ethyl 3–aminobenzoic methanesulfonate salt), buffered with sodium bicarbonate. Upon assured death, the caudal fin was cut with a scalpel and moved to a petri dish with fresh water. After about 30 minutes the parasites started detaching from the fin and moved to the water column, either attaching to the bottom of the petri dish or to the water surface, as described in (Cable et al., 2002). When enough parasites had detached from the fin, experimental fish were anesthetized with a 0.1 mg/ml dose of MS222, buffered with sodium bicarbonate, and their caudal fin was placed near the floating parasites until five parasites attached to the host's fin. The procedure was observed with a dissecting microscope. After infection, experimental fish were revived in fresh water and the number of parasites was counted under a dissecting microscope every 2 or 3 days. In instances where loads were very high but could not perfectly distinguish the precise count, we assigned the value of 200 parasites (for three fish) or 250 (for one fish). Sham-infected fish underwent the same procedure of the experimental infection but no parasites were given to the fish. All control sham-infected focal-fish were also sham-screened every 2 or 3 days. Fish used for the parasite culture were not used for behavioral experiments.

1.2.3 Experimental trials

Experimental trials were carried out in a 170mL swim tunnel respirometer (Loligo® Systems, Viborg, Denmark) filled with aerated and pH-adjusted water and maintained at the temperature of 26.5 $\pm$ 1 °C. All experiments were conducted between January and July 2021, between 8:30 am and 5:30 pm in continuous visible light, as in the fish's holding condition. The swim tunnel respirometer is composed of two glass tubes: a bigger tube for water re-circulation containing a smaller tube, that is the swim and respirometer chamber, where the fish is placed and swims in laminar flow while oxygen consumption is measured (Figure 1.1). Oxygen levels in the respirometer were recorded using

an oxygen dipping probe type PSt3 (PreSens, Regensburg, Germany) and measurements were recorded with Autoresp Software (Loligo® Systems). The water flow in the swim tunnel was calibrated using fluorescence-based Particle Image Velocimetry (PIV). For an illustration of the experimental setup, see Figure 1.1.

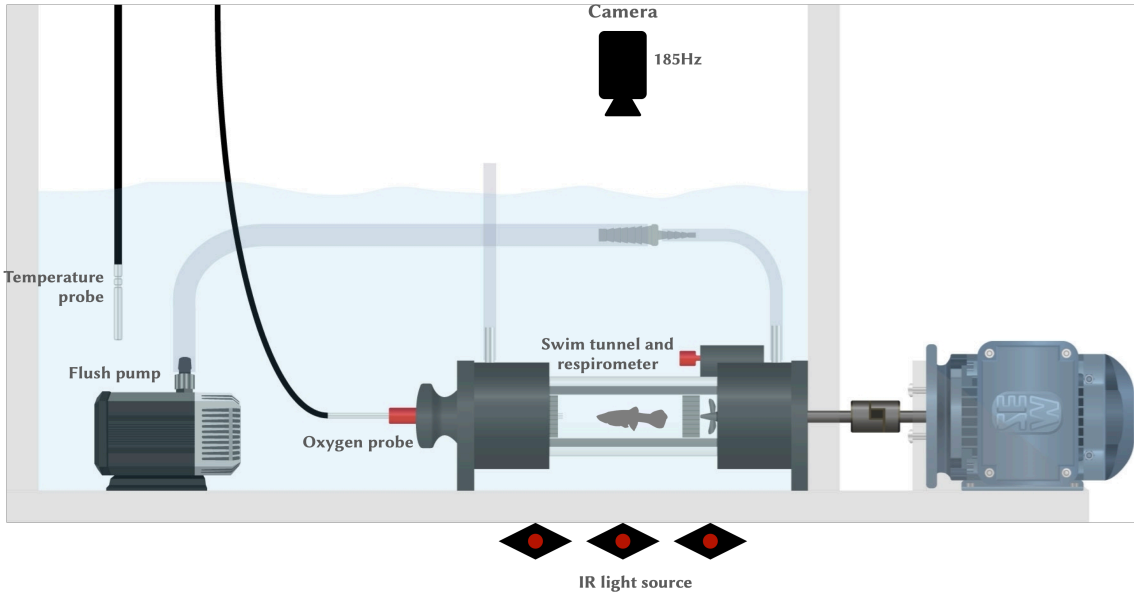


Figure 1.1: **Swim tunnel respirometer tank**. Illustration of the 170mL swim tunnel respirometer tank (Loligo® Systems, Viborg, Denmark), including oxygen dipping probe (red cylinder, left) and motor to control water flow speed (right, propellers). The original image was taken from the Loligo Systems website and adapted to include specifics of our experimental setup. We customized our setup to allow bottom infrared illumination using a custom-built infrared LED array (Winger® IR power LED 850nm 3W). The camera was placed above the tank and videos were recorded at 185Hz.

We measured a total of 63 female guppies. All fish were tested on either day 1 (taken directly from a holding tank), or 5, 10, or 15 days of semi-isolation (that is a condition where the fish is in a single compartment but separated by transparent partitions to allow the fish to see conspecifics in neighboring compartments; see section 1.2.2. Experiments held on condition of Day 1 were excluded from further analysis as fish tested on this day would not be directly comparable to any of the treatment conditions, where fish grow an infection (or sham-infection) in semi-isolation before the experiment. After excluding fish tested on Day 1, we included a total of 52 female guppies, divided between Control (CTR, N=24) and Treatment (TRT, N=28) trials. Both Control and Treatment condition fish were tested in the swim tunnel after either 5, 10, or 15 days of isolation. In the Treatment condition, time in isolation was used to determine either short (5 days), medium (10 days), or long (15 days) infection duration, which, on average, corresponds to low, medium, and high infection loads; however, parasite growth does not always linearly grow with infection duration: parasite growth curve shown in Figure 1.2. In the Control

condition, holding and handling were kept identical to the Treatment, but fish were not exposed to parasite infection. Preliminary analysis suggested a non-monotonic relationship between kinematics and parasite load, therefore in order to simplify the analysis, we categorized parasitized individuals with respect to their parasite load creating low, medium, and high parasite load groups. These groups were assigned according to the 33^{rd} and 66^{th} percentile of the total parasite load distribution, corresponding to 24 and 77 parasites, respectively. This division assigned equal (or nearly equal) numbers in each group (Figure 1.2).

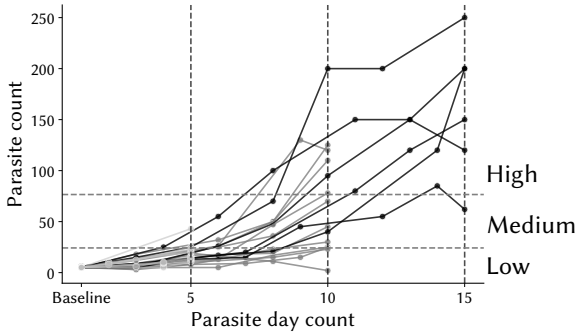


Figure 1.2: **Parasite load growth over time**. Parasite load and its growth from baseline to testing day for all tested fish. Each solid line shows the parasite growth for one fish measured over different days (minimum 5 days, maximum 15 days). The vertical dotted lines highlight the treatment days (5, 10 or 15). *Gyrodactylus* sp. shows approximately an exponential growth over time of a naive infection, followed by a possible decay over the following days. Horizontal dotted lines show the cutoff to determine discrete categories of parasite load, that is low, medium, and high parasite loads. The cutoff is calculated on the entire population of fish, prior to further filtering (N=52). Low cut, 33^{rd} percentile: 24 parasites; high cut, 66^{th} percentile: 77 parasites.

Each fish of the Treatment condition was infected on the first day of the semi-isolation (see more in section 1.2.2) and parasite load was assessed under a dissecting microscope every 2 or 3 days. The same experimental procedure was used for the Control condition, where the Focal-fish was sham-infected. The Control condition was used to verify that the experimental manipulation did not induce an effect on the measured parameters. Each individual's length, width, depth, and mass were measured prior to the trial. Fish were then placed in the respirometer and left to habituate to the chamber for 4 hours at the constant speed of 2 BL·s^{-1}, which corresponded to the lowest flow necessary to ensure stable and constant swimming for our tested population and fish species (Svendsen et al., 2013). The acclimation speed was determined based on data collected with 10 pilot trials. We used a 6.5 minutes loop period with 120s flush, 30s wait, and 240s measurement cycle. After the acclimation time, flow speed was incrementally increased by 0.5 BL·s^{-1} every three loops following a standard U_{crit} protocol (Brett, 1964; Roche et al., 2013a). If a fish touched the comb at the back of the swim tunnel with the caudal

fin, we would encourage swimming by tapping the glass tube. Aquarium plastic plants were also positioned on each side of the swimming chamber to motivate swimming towards the front. Trials were stopped when fish could not swim against the water flow, as determined by the fish resting at the back of the chamber for longer than 10 seconds. The value of U_{crit} is defined using the time the trial stopped, along with the time in that flow, and the loop time (Brett, 1964).

We filtered data after collection based on the evaluation of body condition and identification of outliers defined in the 0.01 percentile of the distribution of maximum oxygen concentration and the reliability R^2 value, which defines the reliability of respirometry measurements. Based on this filtering, a total of 6 fish were excluded for further analysis. One additional trial was excluded for lack of data at each flow speed tested (high rate of unsteady swimming, more details on this calculation below). After this filtering, we obtained a total of 46 trials, divided in N=20 for the CTR and N=26 for the TRT condition. The counting of the day of testing started with the infection (or the sham-infection) procedure, followed by individual housing in tank compartments (see section 1.2.2). For each testing day, we counted N=8 at day 5 (one of which was actually tested at day 6), N=9 at day 10, N=3 at day 15 for the CTR condition, compared to N=7 at day 5, N=11 at day 10 (one of which tested at day 14), N=8 at day 15 (two of which were actually tested at day 16) for the TRT condition. All fish were tested only at one time point and were naive to the experimental setup. In the CTR all the fish are non-parasitized, while in the TRT conditions the focal-fish is infected with a variable number of parasites following the procedure described above 1.2. Top-view swimming was recorded at 185 frames per second for a minimum of 30 seconds at each flow speed using a high-resolution camera (Basler® acA4112-30um) with a 25mm lens. Data acquisition was controlled by video recording software (Motif, loopbio GmbH). The illumination system was custom-built as an array of 15x15 infrared LEDs (Winger® IR power LED 850nm 3W).

1.2.4 Morphology

Before entering the swim tunnel, for each fish, we measured the total length (snout to tail, included), maximum width, maximum depth, and weight. Additional morphological measures were obtained from the video, with the assumption that tracking or video annotation is less sensitive to errors than hand measurement. More specifically, we manually annotated 3 frames per fish to estimate body width from randomly selected video frames (excluding potential frames where body measures would be noisy) and converted the measured lengths to centimeters. Total body length was estimated as the average distance between snout and tail tip across all tracked frames. Standard length was calculated as the mean distance between the snout (S) and the caudal peduncle (T2)

(Figure 1.3).

In addition to this, we calculated two dimensionless numbers, the Scaled Mass Index (SMI) and the Fineness ratio. The SMI is a measure of body condition and is calculated as in Maceda-Veiga et al. (2014) and Peig and Green (2009). The fineness ratio is a measure of body elongation relative to its transverse sectional diameter, and it was calculated based on the code shared on Di Santo et al. (2021) and adapted from Walker et al. (2013).

1.2.5 Physiology

We estimated aerobic metabolic rates indirectly from measurements of oxygen consumption rate (MO_2; mg O_2 kg^{-1} h^{-1}) as a function of swimming speed (U). We calculated the Factorial Aerobic Scope (FAS) as the ratio between MO2 at the end of the experiment, and MO2 at the last three loops at the acclimation speed (2 BL·s^{-1}). FAS is a measure of the proportional increase in oxygen consumption rate above baseline levels and it is commonly considered more robust than absolute aerobic scope measurements (Halsey et al., 2018). The water of the experimental setup was changed at the beginning of each trial, with filtered and pH adjusted water (see section 1.2.1); therefore we did not calculate the background O2 consumption rate.

1.2.6 Kinematics

For each flow speed tested we analyzed top-view videos recorded at 185Hz for a minimum of 30 seconds. We tracked all videos using DeepLabCut (Lauer et al., 2022; Mathis et al., 2018), which allows marker-less posture estimation. Fish were filmed from above the tank and therefore all midline kinematics data are 2-dimensional. We tracked a total of 5 points along the fish midline and 1 point for each pectoral fin. The points along the body were chosen to describe the full curvature of the fish in a simplified way to capture the locomotion patterns of guppies in the different flow regimes. The points were assigned to: snout (S), head (caudal to the brain; H), frontal starting point of the dorsal fin (midbody; M), midpoint between the dorsal fin and the peduncle (approximately 7 scales caudal to M) (tail1; T1), peduncle (tail2; T2), tip of the caudal fin (tailtip; TT), most marginal point of the left pectoral fin (LxP), most marginal point of the right pectoral fin (RxP). An additional point calculated as the average position between H and M was labeled as centroid (C) (Figure 1.3A). For each tracked point of the posture, we exported the raw x, y coordinates of all individuals. To partly account for tracking noise, we smoothed the time series with a moving average with a window size of 5 frames.

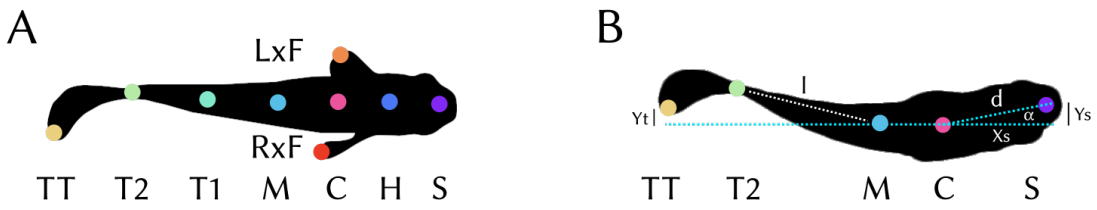


Figure 1.3: **Midline points tracking and analysis.** (A) Videos were analyzed with DeepLabCut (Mathis et al., 2018) after annotating images with 6 points along the body midline, in addition to one point for each pectoral fin. Tracked points are snout (S), head (caudal to the brain; H), frontal starting point of the dorsal fin (midbody; M), midpoint between the dorsal fin and the peduncle (approximately 7 scales caudal to M) (tail1; T1), peduncle (tail2; T2), tip of the caudal fin (tailtip; TT), most marginal point of the left pectoral fin (LxP), most marginal point of the right pectoral fin (RxP). (B) All points were centered according to centroid point (C, pink). After centering to the body centroid C, we determined the oscillation of the points, such as Y_s and Y_t in the figure, corresponding to the y-offset of the snout and the tail, respectively. The heading direction angle α is shown for reference, although the points were not rotated according to α.

Spline and pectoral-fin points at each frame were centered according to the centroid of the body, and all associated metrics are calculated on these transformed points (Figure 1.3B). The heading direction angle (α) was calculated as the arctangent of the horizontal and vertical components of the snout after centering to the centroid (*i.e.* $\arctan(Y_s,X_s)$). The angle α was used for filtering based on heading direction. We used wavelet analysis to calculate the frequency of the tail tip Y_{tt} (Figure 1.3B) (f, as fin-beats x s^{-1} of the tail tip; Hz) and a peak-finding algorithm on the y-offset to calculate tail tip beat amplitude (A). Using the tail tip (TT) frequency and amplitude we calculated the Strouhal number (St) as fA/U, where U is the real speed of the fish (swimming speed detected in the videos and flow speeds). The Strouhal number (St) is a dimensionless parameter that describes efficiency in kinematics across taxa (Taylor et al., 2003). The amplitude of the head oscillation was calculated using a peak-finding algorithm on the y-offset of the snout (Y_s, Figure 1.3B). Using the amplitude values of the head and the tail tip, we then calculated the Head:Tail oscillation amplitude ratio (H:T). The measure of the H:T amplitude ratio has been historically associated to the characterization of locomotion patterns of body and caudal fin (BCF) swimmers, where anguilliform ("eel-like") swimmers should show higher H:T oscillation compared to thunniform ("tuna-like") swimmers, that should show lower head yaw amplitude (Di Santo et al., 2021). For each joint, we measured the phase of the oscillations along the body, but for the body wave estimation, we only took into account values of phase difference between the midbody point M (roughly at the dorsal fin) and the peduncle point T2 (Figure 1.3). As a proxy of body

stiffness, we calculated body wavelength in the body segment M-T2. The body wavelength is a measure of the distance traveled by the wave over one period; moreover, body wavelength is used as a proxy of body stiffness, where low values of propulsive wave should indicate lower body stiffness (or higher body flexibility) (Di Santo et al., 2021). We calculated body wavelength dividing the distance l between the two body points M and T2 (see Figure 1.3B), by the phase difference between the same points, as (l / PD) * 2pi. All kinematic parameters were quantified as a proportion of body length (BL) and were filtered in order to include only steady-swimming kinematics (see more in section 1.2.7. An overview of the kinematic metrics used is found in Figure 1.4.

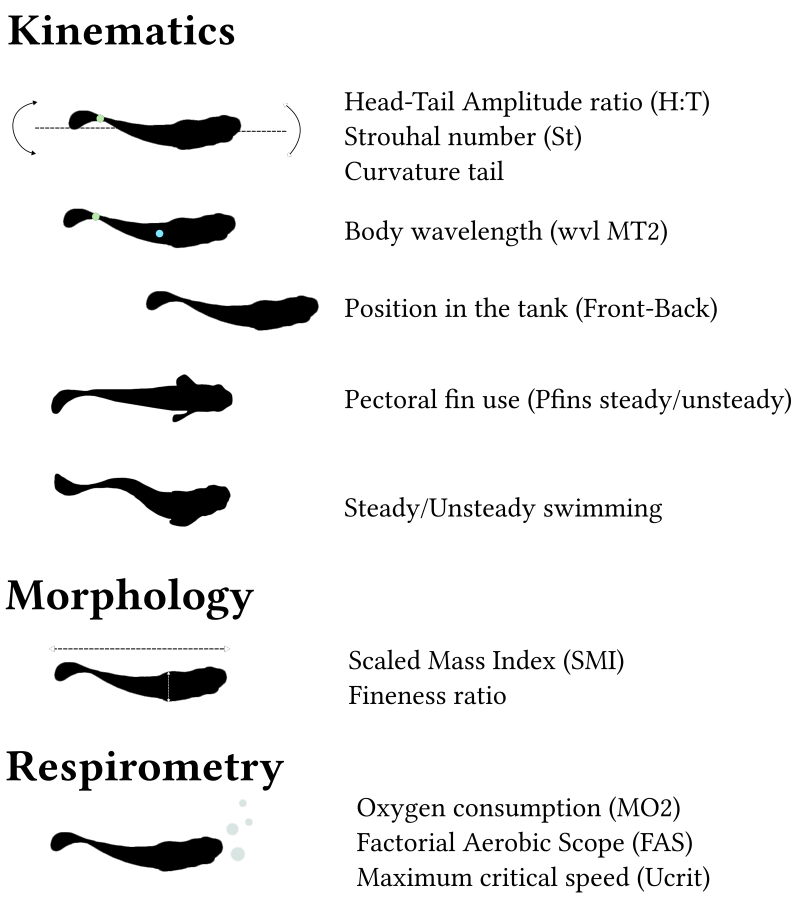


Figure 1.4: **Illustration of the features analyzed.** We analyzed a total of 13 features, including measures of kinematics, physiology, morphology, and performance. We grouped some features and assigned an illustration. Relevant feature abbreviations are written in parentheses next to the name. The kinematic measures include Head:Tail amplitude ratio (H:T), Strouhal number (St), Curvature at the tail tip (Curvature tail), body wavelength (wvl MT2), position in the tank (Front-Back), pectoral fin use in steady and unsteady swimming (Pfins steady/unsteady); physiology measures include oxygen consumption rate (MO2; mg O_2 kg^{-1} h^{-1}) and Factorial Aerobic Scope (FAS); morphology measures include fineness ratio and Scaled Mass Index (SMI); swimming performance measures include maximum critical speed (U_{crit}) and rate of unsteady swimming (Unsteady swimming).

For each pectoral fin, we calculated a percentage of time used, both when in steady and unsteady swimming, which is one of the parameters used for this analysis (more detail in 1.2.7). We estimated pectoral fin movement by calculating the normalized pectoral fin motion speed along the y-offset of the pectoral fins. For calculating the total duration of fin use, we defined windows of minimum consecutive frames of 1/8 of a second, within which the variable of interest (in this case pectoral fin velocity) was above an arbitrary threshold (Figure A.1). The threshold was visually validated on a subset of the videos. When two or more consecutive windows were separated by less than 0.5s, these windows were combined into one. The proportion of time of pectoral fin use was calculated as the number of frames in each window (*i.e.* corresponding to the number of frames where the fish used pectoral fins above the threshold value), divided by the total number of frames for each flow speed. The pectoral fin use was estimated for both steady and unsteady swimming events. In section 1.2.7 we explain how we defined steady and unsteady events.

1.2.7 Steady and Unsteady swimming

We analyzed the entire time series of each recording to quantify spatio-temporal swimming stability. For each flow speed tested, we classified swimming into two categories of steady and unsteady swimming. We defined steady swimming by identifying frames with constant speed, heading orientation, and position in the tank. More specifically, in our definition of steady swimming, the heading direction angle was constrained in $-0.5 < \alpha < 0.5$rad (with 0 rad being head oriented against the water flow); swimming speed was restricted to real speed plus an offset $-0.8 < U_i(t) < 0.8$ BL·s^{-1}; position in the tank was limited to frames where the caudal fin was possibly touching the honeycomb at the back of the swimming chamber. To quantify the proportion of time spent in unsteady swimming, we defined epochs of consecutive frames within which the fish is not steady (see filter above used for pectoral fin use, and see Figure A.2). The minimum unsteady epoch duration length was set to 1/8 of a second. When two or more epochs were found to be separated by less than 0.5s, they were combined into one. Each epoch was padded with 1/4s frames to account for the beginning and the end of an unsteady event. The proportion of time the fish spent unsteady swimming was then calculated as the number of frames found with the filter divided by the total number of recorded frames for that flow speed.

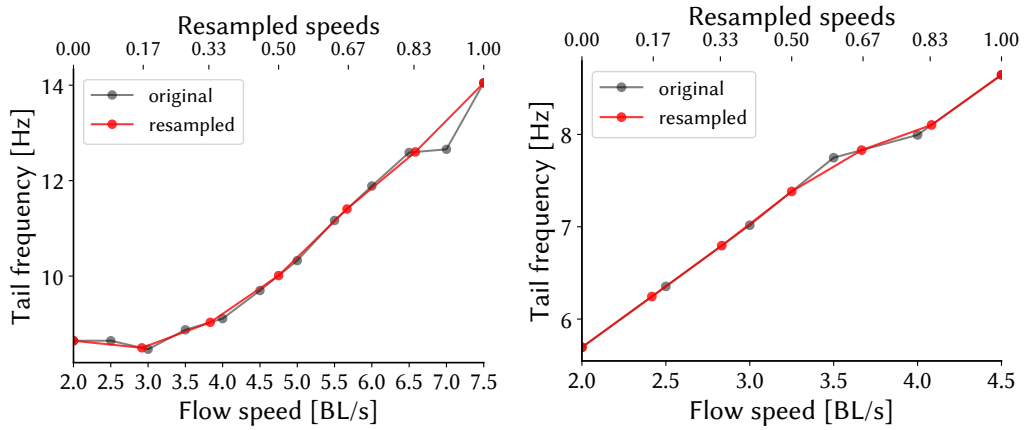
1.2.8 Interpolation of flow speeds

Fish varied in their swimming performance, which we define as their maximum critical speed U_{crit} (Figure A.3). Because of this, we have fewer data points for fish that swim only up to low flow speeds (*e.g.* 5 data points for a fish with $U_{crit} = 4$ BL·s^{-1}), compared to fish that swim up to higher flow speeds (*e.g.* 13 data points for $U_{crit} = 8$ BL·s^{-1}). This discrepancy would bring difficulties in data interpretation due to an inability to draw clear conclusions based on the comparison of kinematics at the same flow speed. For example, if we were to compare kinematics at 4 BL·s^{-1} for a fish close to exhaustion ($U_{crit} = 4$ BL·s^{-1}), to a fish that is at 50% of its performance costs ($U_{crit} = 8$ BL·s^{-1}), this would be an unfair comparison. To account for this, we therefore resampled all the kinematics and MO2 values in order to have the same number of points per fish, fixed to 7 resampled speeds. An example is shown in Figure 1.5, where the values of tail frequency across speeds are compared between two fish: on the left, a fish with measurements up to 7.5 BL·s^{-1} (*i.e.* 12 data points) and thus reduced to 7 points (Figure 1.5a); on the right, a fish with kinematic measurements up to 4.5 BL·s^{-1} (*i.e.* 6 data points) is resampled to increase the points to 7 (Figure 1.5b). The values were then scaled to range between 0 and 1, where 1 represents the maximum swimming speed recorded. In this way, we compared kinematics across fish based on their stage of performance, rather than at the fixed flow speed measure, which could result in individual variation due to differences in physiology, morphology, and so on. All figures showing metrics as a function of speed will therefore refer to the interpolated value of the metric over 7 datapoints (x-axis label: resampled speeds), unless stated otherwise.

1.2.9 LDA

Linear Discriminant Analysis (LDA) is a supervised dimensionality reduction method that uses linear combinations of features to predict the class of a given observation (Xanthopoulos et al., 2013). LDA finds the so-called linear discriminants (LDs, linear combination of predictors) that maximize the separation between classes; these LDs can then be used in combination with other methods to predict the class of individuals. We used the Eigenvalue decomposition solver, with the automatically computed shrinkage of the `sklearn.discriminant_analysis.LinearDiscriminantAnalysis` package (Figure A.4). The results of LDA are shown in Figure 1.11.

Here, using 13 metrics we predict the parasite load class divided into four groups: non-parasitized, low, medium, and high parasite load. The 13 features used include measures of swimming kinematics, morphology, and physiology. The kinematic features are: Head:Tail amplitude ratio (H:T), Strouhal number (St), Curvature at the tail



(a) Interpolation from more data points to less (b) Interpolation from fewer data points to more

Figure 1.5: **Interpolation of metrics across flow speed.** Left: an example of a reduction of data points for an individual with more data points than 7 (*i.e.* U_{crit} value higher than 5 BL·s^{-1}, spacing of 0.5 BL·s^{-1} between data points). In this example, 12 points were reduced to 7. Right: an example of an increase of data points for an individual with less data points than 7 (*i.e.* U_{crit} value lower than 5 BL·s^{-1}). In this example, 6 points were increased to 7. The top axis shows the scaling of the resampled data points, ranging from 0 to 1, where 1 corresponds to the highest recorded speed of the fish. The gray line shows the original data points of an example metric (in this case Tail Frequency) across flow speeds. The red line shows the resampled data over 7 points.

tip (Curvature tail), body wavelength between M (beginning of dorsal fin) and T2 (caudal peduncle) (wvl T2), pectoral fin use when steady and unsteady (Pfins steady, Pfins unsteady), position in the swim tunnel (Front-Back), and rate of unsteady swimming (Unsteady). The morphology features are: Scaled Mass Index (SMI) and the Fineness ratio. The respirometry features are: oxygen consumption rate (MO2) and Factorial Aerobic Scope (FAS). The performance feature is the maximum critical speed (U_{crit}). For an illustration of the features and the abbreviations see Figure 1.4. Using these metrics we created a data matrix M_{ij}, where each row i=1...46 represents one fish, and columns j=1...91 are the 13 features times the 7 resampled flow speeds. We normalized the data matrix **M** by subtracting the mean and dividing by the standard deviation. On this data matrix, we performed both LDA and principal component analysis (see next paragraph). To calculate the model accuracy we used a classifier based on the euclidean distance of each point from the clusters' centers. We evaluated the accuracy using the first two LDs.

1.2.10 PCA and hierarchical clustering

Principal component analysis is an unsupervised dimensionality reduction method commonly used to discover underlying patterns across features (Xanthopoulos et al., 2013). The result of PCA is a matrix U_{ij}, where i represents the fish i and j = 1...91 for the PCA components. We additionally performed Ward hierarchical clustering, implemented in

Python in the package `scipy.cluster.hierarchy`, to obtain the results shown in Figure 1.12. Ward clustering minimizes the overall within-cluster variance.

1.2.11 Statistical analysis

Descriptive statistics were computed for morphology and respirometry variables and mean values were compared across parasite load groups using a one-way ANOVA followed by a Tukey Honestly Significant test (α = 0.05) and multiple comparison correction.

Correlation values were calculated using Pearson correlation coefficient. All values are reported with the correlation coefficient and the p-value.

1.3 Results

We analyzed a total of N=46 fish divided into non-parasitized (N=20) and parasitized fish (total N=26), then divided for low (N=7), medium (N=9), high (N=10) parasite loads. Videos were recorded at 185 frames per second for a minimum of 30 seconds for each flow speed, and each fish was tested at increasing flow speeds until exhaustion. For each fish, we measured a number of kinematics, physiology, and morphology metrics. A representation of the metrics used in the analysis is found in Figure 1.4 (see also 1.2.6).

1.3.1 Morphology

Morphology measures included body length, standard length, weight, depth, width, fineness ratio, and Scaled Mass Index (SMI). Body length was estimated from the posture tracking. For the tested fish, we obtained an average body length of 3 $\pm$ 0.16 cm, standard length of 2.41 $\pm$ 0.13 cm, weight of 0.3 $\pm$ 0.1 gr, depth (measured at the deepest point of the body) of 0.46 $\pm$ 0.05 cm, body width (estimated manually the videos) of 0.47 $\pm$ 0.06 cm, fineness ratio of 5.59 $\pm$ 0.46 and SMI of 0.29 $\pm$ 0.04. None of these values revealed significant differences across treatments, that is between non-parasitized individuals, low, medium, and high parasite loads (one-way ANOVA). Figure 1.6 shows values for body length, SMI, and fineness ratio, grouped by parasite loads. SMI is generally used as an index of body condition where a larger SMI indicates shorter body lengths but larger weight and it positively correlates with body width. Moreover, SMI is also positively correlated with U_{crit} (r=0.37, p=0.01), which is explained by the fact that larger fish (longer body length) have lower relative critical speeds (in body lengths). However, the correlation between SMI and U_{crit} is more pronounced for parasitized fish, compared to non-parasitized fish. More specifically, we find no correlation between SMI and U_{crit} for

non-parasitized fish (r = 0.09, p>0.05), compared to a stronger correlation for parasitized fish (r = 0.54, p=0.004) (Figure A.5a). We find a similar (but opposite in sign) relationship between Fineness ratio and U_{crit}, with correlation values of r=0.16 and r=-0.31 (p>0.05) for non-parasitized and parasitized fish groups, respectively (Figure A.5b).

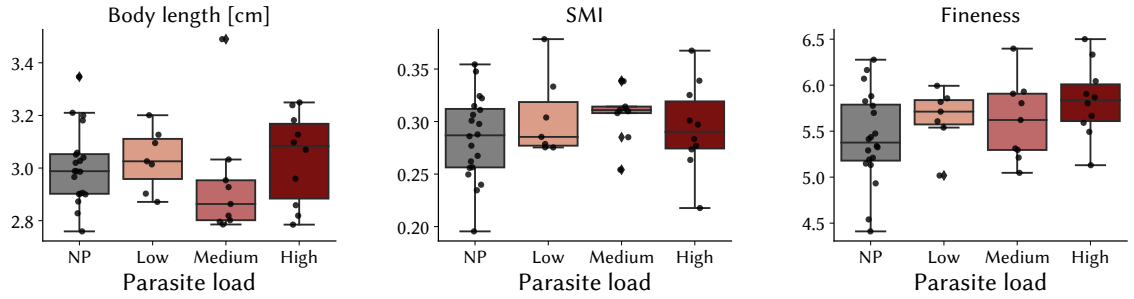


Figure 1.6: **Morphology measures across parasite loads.** Morphology metrics of body length, SMI, and fineness ratio (from left to right). None of these values were statistically different across groups (one-way ANOVA). Box plots give the median, interquartile range (box), and values that sit outside of 1.5 times the interquartile range (whiskers).

1.3.2 Oxygen consumption and Fractional Aerobic Scope

We collected oxygen consumption (MO2; mg O_2 kg^{-1} h^{-1}) and Fractional Aerobic Scope (FAS) data for each fish tested, at each flow speed. MO2 and FAS values did significantly differ across treatments (one-way ANOVA, p=0.0388 and p=0.006, respectively). More specifically, for the average MO2 values, low-parasite load individuals showed significantly lower MO2 values than high-parasite load individuals (post-hoc Tukey Honestly Significant Difference p=0.0021, Bonferroni corrected). On the other hand, FAS showed a significant difference between the non-parasitized individuals and the low-load group (post-hoc Tukey Honestly Significant Difference p=0.0014). Among the parasitized individuals, however, FAS is negatively correlated with increasing parasite load (Figure A.6).

Figure 1.7 shows the correlation between a subset of morphology and physiology measures and U_{crit}. Correlation values lower than 0.3 are not shown. Average MO2 values positively correlate with fineness ratio (r=0.37,p=0.01) and negatively correlate with SMI (r=-0.37, p=0.01). This suggests higher MO2 values for relatively smaller mass bodies (width and weight). A more extended version of the correlation matrix that includes more variables as well as kinematics information is found in the Appendix in Figure A.7.

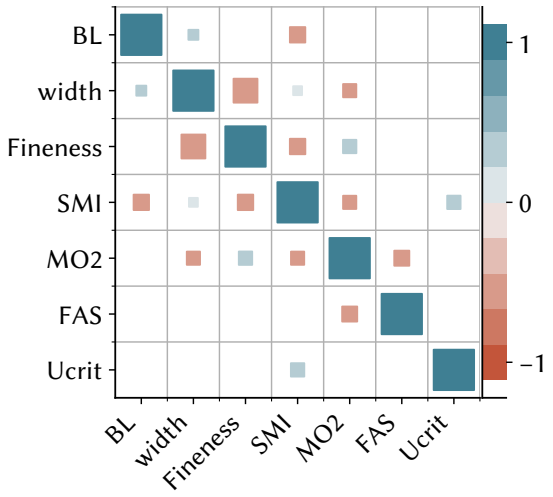


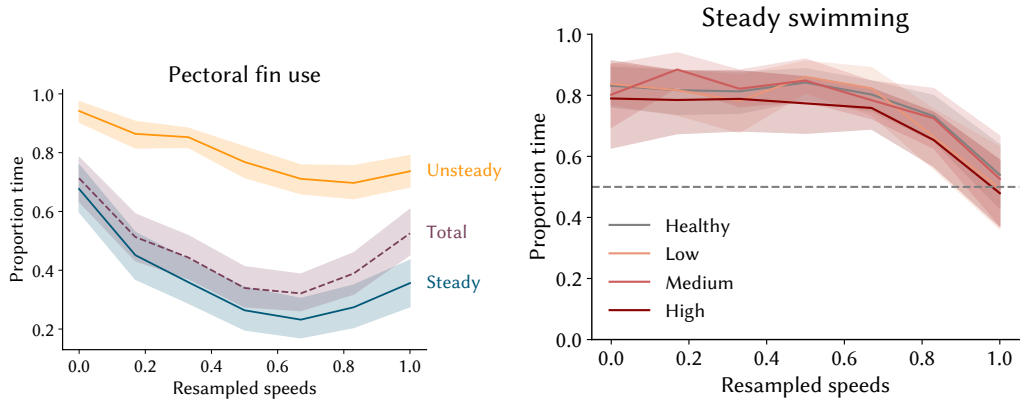
Figure 1.7: **Correlation morphology and physiology.** Correlation matrix of a subset of the morphology and physiology measures. Morphology and physiology measures only partly correlate. Body width and SMI negatively correlate with average MO2 (r=-0.37, p=0.01), whereas fineness ratio positively correlates with average MO2 (r=0.37,p=0.01). The size of the colored square represents the magnitude of the Pearson correlation coefficient. Only correlations higher than r=0.3 are shown in the figure.

1.3.3 Pectoral fin use

It has been previously described that guppies mainly use pectoral fins use at low flow speeds (Svendsen et al., 2013). With our data we confirm this finding trend, showing an average decrease in pectoral fin use with increasing flow speeds (Figure 1.8a). However, Figure 1.8a also shows that the average steady-swimming pectoral fin use (blue line) decreases with increasing flow speeds, while the pectoral fin use during unsteady swimming (orange line) is present at high rates at all flow speeds. On average pectoral fins are used 37% of the time during steady-swimming, compared to 79% of the time when during unsteady swimming (averaged across swimming speeds) (Figure 1.8a). We do not find a difference in the pectoral fin use rate (both steady and unsteady) across parasite load groups (Figure A.8)

1.3.4 General swimming performance

Our definition of swimming performance refers to the maximum critical swimming speed at exhaustion (*i.e.* U_{crit}, speed of the fish at the time when it stops swimming against the water flow, see section 1.2.3). In the evaluation of swimming performance, we also include the ability of a fish to swim steadily in increasing flow regimes, which is commonly used to assess possible trade-offs between performance and other behavioral traits. Here, we first describe the variation in U_{crit} according to their parasite loads,



(a) Pectoral fin use

(b) Steady swimming rate

Figure 1.8: **Pectoral fin use and Steady Swimming rates.** (a) Pectoral fin use during steady and unsteady swimming and total pectoral fin use as a function of flow speed. The average steady-swimming pectoral fin use (blue line) decreases with increasing flow speeds. The pectoral fin used during unsteady swimming (orange line) is present at high rates at all flow speeds. On average pectoral fins are used 37% of the time during steady-swimming, compared to 79% of the time when during unsteady swimming (averaged across swimming speeds). (b) Steady swimming rate as a function of flow speed and parasite load. We do not find a difference in the rate of fin use between non-parasitized fish, and low, medium, and high parasite load fish. Line plots show the mean and 95% CI of the mean.

followed by a description of the proportion of steady versus unsteady swimming.

Fish in the population vary in their maximum critical speed (Figure A.3). We measure an average U_{crit} of 5.9 $\pm$ 0.8 BL. We find no difference of U_{crit} among treatments (more details in the following sections, Figure 1.10B, U_{crit} BL). Overall, we also do not find any significant correlation between U_{crit} and physiology measures (Figure 1.7). As previously shown in section 1.3.1, we find that SMI is positively correlated with U_{crit}, which is explained by shorter fish reaching relatively higher swimming speeds in units of BL·s^{-1}(Figure 1.7 and Figure A.5a).

In addition to U_{crit}, we analyzed rates of steady and unsteady swimming which alternate at various ratios across flow regimes. At lower swimming speeds, steady swimming is the prevalent mode, characterizing, on average across fish, 82% of the swimming time (Figure 1.8b). Fish reach exhaustion when unsteady swimming rate is, on average across fish, 51% of the swimming time. When we compare the ratio of steady and unsteady swimming within parasite load classes we don't find a difference among treatments (Figure 1.8b).

1.3.5 Pectoral fin use, body stiffness and swimming performance

It was previously shown that pectoral fin use in guppies is associated with increased energy expenditure and decreased U_{crit} (Killen et al., 2016; Svendsen et al., 2013). To

further investigate this relationship, we compare the pectoral fin use rate as a function of discrete swimming performance groups. To this end, we separate our fish population into four categories of fish swimmers: 'bad', 'medium', 'good', and 'best, as defined by the 25th, 50th, and 75th quantiles of all recorded U_{crit} values. We find that the best performers show a constant decrease in pectoral fin use compared to the bad performers (Figure 1.9). In general, the best performers use pectoral fins in steady swimming conditions at low flow speeds, with an exponential decrease in use at higher flow speeds. By contrast, bad performers do not simply decrease pectoral fin use with increasing flow speeds but rather show an initial decrease with increasing flow speeds, but then an increase at the highest flow speed reached (Figure 1.9). Similarly, we compare the rate of unsteady swimming for the 'bad' and 'best' swimmers. In this case, we find that the 'bad' performers show an early increase of 50% unsteady swimming at about 5 BL·s^{-1}, compared to the 'best' performers that reach the same rate of unsteady swimming at about 7 BL·s^{-1}.

Another common predictor of swimming performance is body undulation and stiffness, for example as used to distinguish more classic swimming modes across species (Di Santo et al., 2021). Our data show a clear correlation of U_{crit} with values of H:T amplitude ratio. More specifically, values of H:T amplitude ratio at low flow speeds negatively correlate with U_{crit} values (r=-0.3,p=0.04 at lowest speed), but the correlation decreases with increasing flow speeds (Figure A.9A, B). In addition to this, we also investigated the difference in Head:Tail amplitude ratio for the discrete categories of swimming performers. Figure 1.9 shows how 'bad' performers also have higher values of H:T amplitude ratio, as opposed to the 'best' performers, which for all flow speeds show lower H:T amplitude ratio values. As a proxy of body stiffness, we used values of body wavelength; however, we did not find any correlation between body wavelength and U_{crit}. All the kinematic features as a function of the performer group can be found in the Appendix in Figure A.10 and Figure A.11.

1.3.6 Descriptive features

We selected a total of thirteen features to describe differences across the treatment groups of non-parasitized, low, medium, and high values of parasite load (NP, L, M, H). Figure 1.10 shows the individual feature distributions for each fish of all parasite load groups. Figure 1.10a shows features that count one point per-fish and per-resampled-flow speed tested, while 1.10b shows per-trial features (*i.e.* features with one value measured per fish). Non-parasitized fish record low FAS and relatively lower Fineness ratio than average (Figure 1.10b, see also oxygen consumption section), as well as relatively lower Head:Tail (H:T) amplitude ratio and higher curvature at the tail (Figure 1.10a).

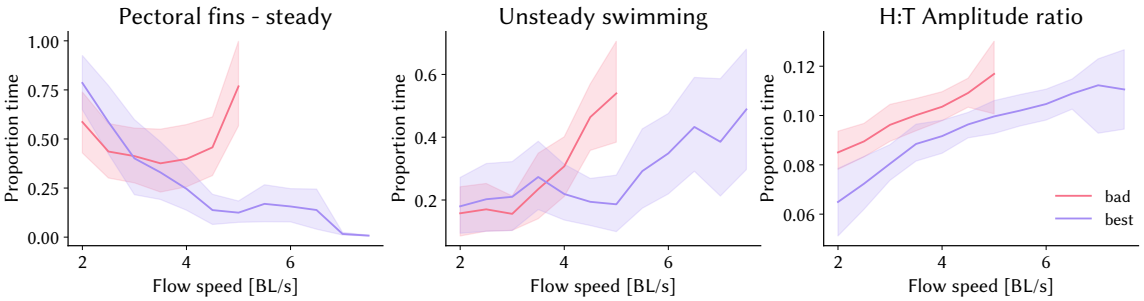


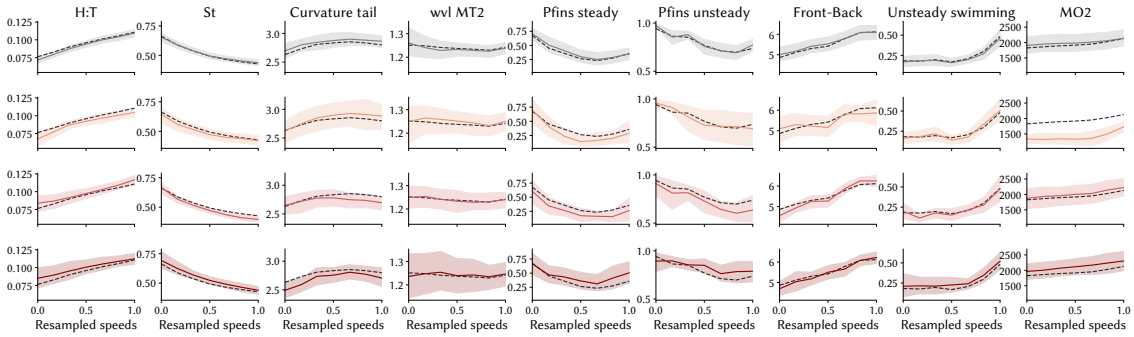
Figure 1.9: **Pectoral fin use and unsteady swimming as a function of swimming performance.** Comparison of 'bad' and 'best' swimming performers on metrics of pectoral fin use during steady swimming (left), unsteady swimming (middle), and Head:Tail (H:T) amplitude ratio (right). The best performers show a constant decrease in pectoral fin use compared to the bad performers (left plot, Pectoral fins). Bad performers show an early increase rate of unsteady swimming compared to the best performers (middle, Unsteady swimming). Bad performers also have higher values of H:T amplitude ratio, as opposed to the best performers, which show lower H:T amplitude ratio values at all flow speeds (right plot, H:T amplitude ratio). Line plots show the mean and 95% CI of the mean.

Low-parasite load fish mainly show lower H:T and Strouhal numbers than average, low MO2, high FAS, and SMI (Figure 1.10). Medium-parasite load fish on average show relatively higher H:T and pectoral fin use at steady swimming (Figure 1.10), in addition to relatively higher SMI, and FAS (Figure 1.10b). Finally, high-parasite load fish are mainly characterized by higher values of H:T, Strouhal number, and pectoral fin use for steady swimming at high flow speeds than the mean population, lower tail curvature, as well as higher MO2 and Fineness ratio (Figure 1.10).

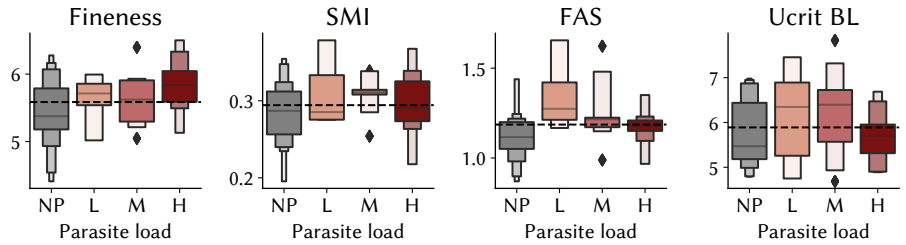
1.3.7 LDA

We used Linear Discriminant Analysis (LDA) to compute the axes that maximize separation between classes of parasite load in the data using the 13 features previously described (Section 1.3.6, Figure 1.4). All metrics, except for Fineness, SMI, FAS, and U_{crit} were computed as a function of flow speed. LDA is a dimensionality reduction method that by contrast to PCA is 'supervised' in that it takes into account class-labels to maximize the ratio of inter-subject to intra-subject variability. Before performing LDA, all data are normalized by subtracting the mean and dividing by the standard deviation of each column. LDA is calculated on a data matrix Mij where each row i is for a single fish and the columns j is for each average behavioral metric at each resampled flow speed. We used LDA to ask what features most differentiate between groups of parasite loads (previously described in the Methods 1.2.2). Looking at the first two components, the model reaches an accuracy of 93%.

Figure 1.11A shows the first two Linear Discriminants (LDs) and the ellipses show



(a) Features as a function of flow speed



(b) Per-trial features

Figure 1.10: **LDA input features.** (a) Line plots for each LDA input feature (columns) are shown as a function of resampled flow speed. Each row represents a parasite load group, from top to bottom: non-parasitized, low, medium, and high parasite load. The black line shows the average calculated across all the individuals in the population. Line plots show the mean and 95% CI of the mean. (b) Letter-value plots show the distribution of four LDA input features that are measured one time in the trial. These features are Fineness ratio, SMI, FAS, and U_{crit}. The dotted line shows the mean across all the individuals in the population. Features are displayed as a function of parasite load group (x-axis). List of abbreviations: H:T: Heat:Tail amplitude ratio; St: Strouhal number; Curvature tail: Curvature measured at the tail tip TT; wvl MT2: body wavelength between M (dorsal fin) and T2 (peduncle); Pfins steady/unsteady: pectoral fin use in steady and unsteady swimming; Front-Back: relative position in the tank; Unsteady swimming: rate of unsteady swimming, including bursting; MO2: oxygen consumption rate; FAS: Factorial Aerobic Scope; SMI: Scaled Mass Index (SMI); U_{crit} BL: maximum critical speed (in body lengths). For a list of the features used also see Figure 1.4.

the 95% confidence interval around each category. We find a general separation between non-parasitized (NP, gray) and parasitized fish of all parasite load categories (colored ellipses), which is mainly captured by the first linear discriminant. In fact, we see that LD1 mainly separates NP points (left on the x-axis, Figure 1.11), from low, medium, and high parasite load groups, whereas LD2 in particular captures the separation between low and high parasite load groups (right on the x-axis, Figure 1.11). All the weightings of each metric at each flow speed, for the first four principal components are shown in the Appendix in Figure A.12. To simplify the resulting interpretation we computed the average weighting for each feature across flow speeds. The feature weightings averaged across interpolated flow speed are shown in Figure 1.11B, where positive/negative

weightings in the LDA components represent higher/lower values of a metric with respect to the trial average. The average weighting direction of each feature is shown in Figure 1.11A. The direction of the vector indicates the contribution of each metric. The angles between vectors encode the correlation of the metrics in this space, where small angles represent a high positive correlation, right angles represent a lack of correlation, and opposite angles represent a high negative correlation. The magnitude of the vector is scaled for plotting purposes, the actual weightings of each feature averaged across flow speed are shown in 1.11B.

There is a substantial overlap among kinematics, morphology, and respirometry features in the dimensions computed with the LDA. However, the main contributing features of the first LD are values of respirometry and morphology (FAS, Fineness), followed by kinematics with H:T amplitude ratio, and then again morphology with SMI (Figure A.13). These are therefore the first four main features that characterize a group of non-parasitized fish as opposed to the different classes of parasitized fish. Similarly, the second LD is mainly separating low from high parasite load. The most important feature is tail curvature, followed by H:T, MO2, and St (Figure A.13).

Non-parasitized fish mainly occupy space that is negative on the x-axis (Figure 1.11A), and these reflect low FAS and relatively lower Fineness ratio (Figure 1.10B, see section 1.3.2), as well as relatively lower H:T and higher curvature at the tail (Figure 1.10a). By contrast to non-parasitized fish and positive on the x-axis, we find low-, medium- and high-parasite load fish groups. The characterizing features of these points are higher FAS, fineness, H:T, and SMI (Figure 1.11A and Figure A.13). The low-parasite load fish show points only positive on LD2, whereas high-parasite load fish mainly have points negative on LD2. Low and high-parasite load fish generally show opposite kinematic values (opposite sign on LD2) (Figure 1.11). For more descriptives of the features, see section 1.3.6.

Overall, non-parasitized fish are distinguished from parasitized fish mainly by morphology and physiology measures, with some contribution of values of H:T amplitude ratio. On average, non-parasitized fish, compared to parasitized fish, have smaller FAS, together with smaller SMI (longer and skinnier bodies), lower Fineness ratio (more rounded), and lower H:T amplitude ratio. Based on the LDA results we find that with the features we include, we have enough discriminative power to distinguish groups of non-parasitized, low, medium, and high parasite load fish groups.

1.3.8 PCA

Using the same 13 features as previously used for the LDA (full list in section 1.2.9 and Figure 1.4), we used PCA and clustering to identify patterns and similarities in behavior

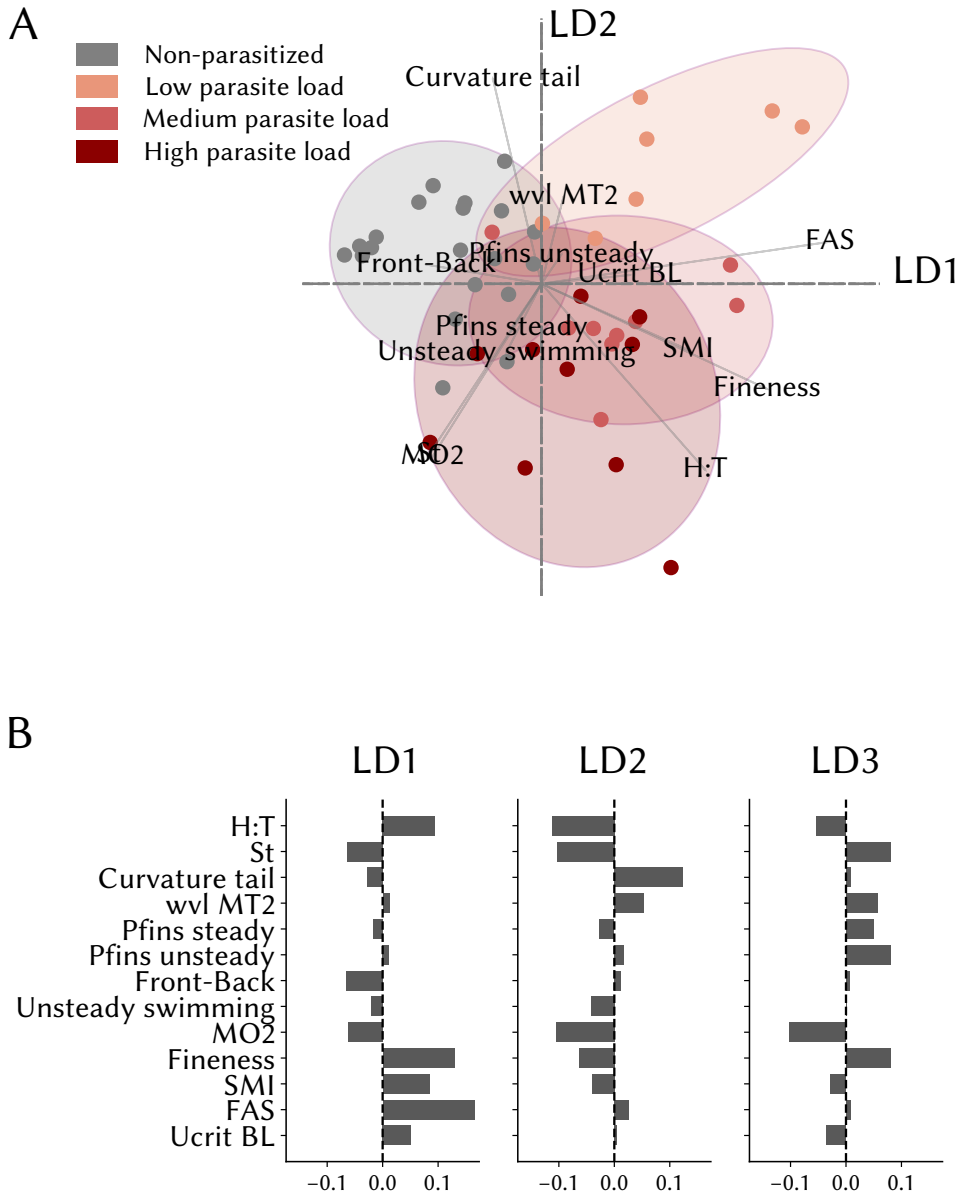


Figure 1.11: **LDA results.** (A) First two Linear Discriminants (LDs) that include 13 metrics. Each color represents a different treatment, that is non-parasitized, low, medium, and high parasite load individuals. The ellipses show the 95% confidence interval around each category. The overlaid weightings show the relative contribution and sign of each metric (averaged across flow speed), also shown in (B). (B) Weightings of all the features for the first three LDs. These weightings have been averaged across flow speed, the complete set of weightings for each metric and flow speed can be found in Figure A.12. For a list of the abbreviations, see Figure 1.4

using a combination of morphology, physiology and kinematic measures. All metrics, except for Fineness, SMI, FAS, and U_{crit} were computed as a function of flow speed (these have one point per trial). A representation of the metrics used can be found in Figure 1.4.

While with LDA the class labels are used to find the combinations of metrics that most separate data points in each class, PCA does not take class labels as input. Instead, using PCA we find the weightings of each metric that explain the largest percentage variance in the data. Before performing PCA, all data are normalized by subtracting the mean and dividing by the standard deviation of each column. The PCA decomposition is calculated on the same data matrix M_{ij} used for LDA, where each row *i* is for a single fish and the columns *j* is for each average behavioral metric at each resampled flow speed. All weightings of each metric at each flow speed, for the first four principal components, are shown in the Appendix in Figure A.14. For simplicity, we computed the average weighting for each feature across flow speeds. The feature weightings averaged across interpolated flow speed are shown in Figure 1.12A, where positive/negative weightings in the PCA components represent higher/lower values of a metric with respect to the trial average. We find that the first principal component explains the largest variance (19.5%), mainly weighted by respirometry and morphology. The second PCA component explains 13.7% and is mainly weighted by kinematic metrics such as curvature of the tail, Strouhal number, and body wavelength. The third PCA component explains 11.9% of the total variance and is most strongly weighted by U_{crit}, FAS, body wavelength, and Head:Tail amplitude ratio. Finally, the fourth PCA component (10% of the total variance) is mainly weighted by the Fineness ratio, U_{crit}, and body wavelength. In total, the first four PCA components explain 55.1% of the total variance of all the features included (Figure 1.12A).

In order to identify main swimming modes, we applied Ward hierarchical clustering (Smith et al., 2021; Ward, 1963). The dendrogram structure built after hierarchical clustering is visible in Figure 1.12B, where each node in the cluster contains a group of similar features. We used three of the clusters found (dotted lines in Figure 1.12B) to describe important aspects of the variation in our data, with the goal of describing trends in the swimming modes that we observed and quantified. The decision regarding the number of clusters is arbitrary, but can still is still valuable to aid data visualization and quantification of the trends that emerge with respect to the combination of features. With this in mind, we chose three clusters to describe dominant swimming modes, each represented by differences predominantly in the first two PCA components (Figure 1.12C).

Each data point in Figure 1.12C shows an individual fish projected onto the first two PCA modes; the color of the points indicates the cluster number. Overlayed on the points

are the loadings of each feature averaged across flow speeds. As in Figure 1.11, the direction of the vector indicates the contribution of each metric to the principal component space. The angles between vectors encode the correlation of the metrics in this space, where small angles represent a high positive correlation, right angles represent a lack of correlation and opposite angles represent a high negative correlation. Each cluster is described by a combination of the features used, characterizing the individuals within that group. The features describing each cluster are shown in Figure 1.12D, for all features that have been evaluated over flow speeds, and in Figure 1.12E for those features that have one value per trial.

One aspect that is revealed with the decomposition is that the clusters are well separated by differences in swimming performance, in this case, U_{crit} values. In fact, Figure 1.12E, bottom right, shows how two clusters are described by relatively lower U_{crit} values than the mean (Cluster 1 and 2), compared to Cluster 3 with higher U_{crit} values. This separation is similar to what we previously described in the section 1.3.5, although in this case, we do not use quantiles of U_{crit} to split the dataset, but PCA followed by clustering. In a simplified interpretation, Cluster 1 and 2 include fish with low U_{crit}, while Cluster 3 includes good performers, as in fish with high U_{crit}.

The main difference between Cluster 1 and 2 is that Cluster 1 includes fish with a specific morphology with a higher fineness ratio and lower SMI, as well as a higher rate of pectoral fin use at higher flow speeds, compared to Cluster 2 that is possibly described by more inefficient kinematics (higher St, higher Curvature). More specifically, we now detail here all the characterizing features for each cluster. Cluster 1 (Blue) includes N=10 fish and it includes individuals that on average have lower U_{crit} values than the population average. All points of Cluster 1 are positive on PC1, and all but three points are positive on PC2 (Figure 1.12C); therefore the main descriptors of this cluster are morphology and respirometry metrics (PC1), in addition to some kinematics (PC2). More specifically, fish belonging to Cluster 1, differently from Cluster 2 (see later), exhibit a clear morphological signature, with higher Fineness ratio values and lower SMI (Figure 1.12E). This means that on average fish belonging to Cluster 1 are larger in body length and are thinner (low SMI), and and more elongated (high Fineness). For the respirometry values, these fish also show low FAS together with larger MO2 (Figure 1.12D, E). Finally, when we look at the kinematics, we find that fish of Cluster 1 are mainly characterized by lower St at low flow speeds, low curvature of the tail, high rate of pectoral fin use for steady swimming at high flow speed (reminding of the 'bad' performers in section 1.3.5), and relatively more frontal swimming position at low flow speeds (Figure 1.12D, E, Cluster 1 blue). Moreover, of the fish of Cluster 1, 5 out of 10 fish are parasitized and 5 are indicated as being in 'bad condition' (2 of which are non-parasitized) (more in the

Discussion).

Cluster 2 (orange) includes N=15 fish, also in this case including individuals with lower U_{crit} values than average. Most points of Cluster 2 are negative on PC2 (Figure 1.12C), which means that these are best described by some kinematic features (Figure 1.12A). In fact, this group shows a clear trend with Strouhal number and Curvature of the tail being higher than the mean. This suggests a possible decreased swimming efficiency, which might be reflected in poorer swimming performance. The same fish also have low body wavelength values (more pronounced at low flow speeds), which indicates more flexible (or less stiff) bodies. Finally, the same Cluster additionally shows higher pectoral fin use during unsteady swimming at high flow speeds, more posterior swimming positions in the tunnel, and relatively lower MO2 values (which might be explained by the fact that they reach lower critical speeds) (Figure 1.12D, E, Cluster 2 orange). Of Cluster 2, 7 out of 15 fish are parasitized, 6 are marked as being in 'bad condition' (5 of which are non-parasitized), and 3 (out of 5 in total) females are marked as pregnant (more in the Discussion).

Finally, Cluster 3 (Green) includes N=21 fish and is mainly described by high U_{crit} values. Individuals in Cluster 3 are generally negative on PC1 and positive on PC2 (Figure 1.12C), which suggests that fish in this cluster have higher SMI, FAS, and U_{crit}, as well as lower MO2 values. Besides morphology and respirometry, Cluster 3 is also characterized by a relatively lower Strouhal number at high flow speeds, higher body wavelength at high flow speeds, and lower pectoral fin use (stead/unsteady) at higher flow speeds (Figure 1.12D, E, Cluster 3 green). In Cluster 3, 14 out of 21 fish are parasitized. The same cluster also includes 2 of the 5 pregnant females, and 4 of the 15 fish marked for 'bad conditions' (1 of which is non-parasitized). Taken together, these results show that our analysis provides a reliable indication of features that are representative of some subpopulations of the tested individuals, possibly characterizing the swimming modes of the tested population.

1.4 Discussion

In our study, we combine high-temporal resolution kinematics analysis with respirometry measurements to better understand the relationship between swimming performance and parasitism. We characterize intraspecific variation in guppies with different parasite loads, tested in a swim tunnel at increasingly faster water flow speeds. Using LDA on metrics of morphology, physiology, kinematics, and performance, we differentiate between treatments of parasite loads, although with some overlap across classes. We further investigated the general variation in the population by using PCA, with which we

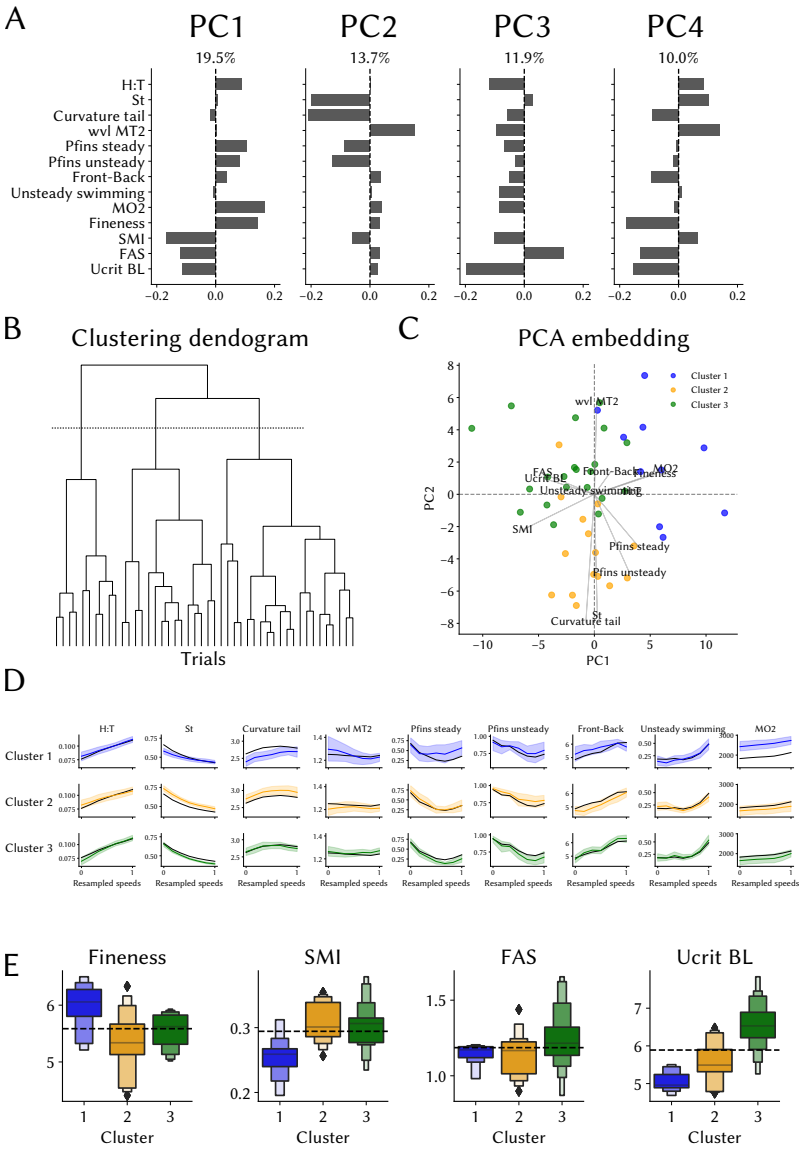


Figure 1.12: **PCA results.** (A) Feature weightings averaged across interpolated flow speed of the first four principal components. Positive/negative weightings in the PCA components represent higher/lower values of a metric with respect to the trial average. The four PCA components explain 20.7%, 14.7%, 11.3%, and 8.4% of the total variance, respectively. (B) Clustering dendrogram. (C) Individual fish projected onto the first two PCA modes. Each point represents one fish; the colors correspond to a 3-cluster division, identified via Ward hierarchical clustering. Overlayed on the points we included the weightings of each feature averaged across flow speeds. (D) Distribution of eight kinematic values and one respirometry (MO2; mg O_2 kg^{-1} h^{-1}) feature as a function of resampled flow speeds. Values are separated by Cluster (Cluster 1 first row, blue color; Cluter 2 second row, orange color; Cluster 3 third row, green color). The black line shows the average calculated across all the individuals in the population. (E) Distribution features with one point per trial (thus not measured as a function of speed), divided by cluster number (x-axis); these are Fineness ratio, SMI, FAS, and U_{crit}. The dotted line shows the mean across all the individuals in the population. For a list of the abbreviations, see Figure 1.4.

describe different swimming modes defined by combinations of features. We find that performance as measured by U_{crit} is a good predictor of different swimming kinematics, and we further investigated the general patterns of variation. This revealed different swimming "styles", which we describe as: fast swimmers with pectoral fin modulation over flow speeds; slow swimmers with inefficient kinematics and high pectoral fin use; slow swimmers with longer/thinner bodies. Using this generalized approach, the results confirm those of Figure 1.9, where we find that the differences in some kinematic features (for example Head:tail amplitude ratio, Unsteady swimming, pectoral fin use) correspond to different swimming performances (Figure 1.12). These results suggest that by using features of morphology, physiology, kinematics, and performance, we can characterize phenotypes in the population and possibly distinguish non-parasitized and fish infected with different loads of parasites.

In this study, we do not find a linear relationship between respirometry values and individuals' parasite load. Instead, we find that variations in physiology measurements are better described in combination with metrics of swimming kinematics. For example, fish with low parasite loads measure the highest Factorial Aerobic Scope (FAS) and the lowest oxygen consumption rates (MO2) values, compared to the rest of the population. By contrast, fish with high-parasite loads measure, on average, low FAS values and higher MO2 compared to the low-parasite load group; however, the FAS and MO2 values of the high-parasite load group are similar to those of non-parasitized fish. Therefore, it would be hard to distinguish high-parasite load fish from non-parasitized fish based on FAS and MO2 values only, without also considering morphology and kinematic differences, as detailed below. Host-parasite interaction is multidimensional and complex (McElroy and de Buron, 2014; Thomas et al., 2010; Timi and Poulin, 2020) and in some cases, the outcome of an infection will depend on several levels. Moreover, it is sometimes difficult to distinguish between direct and indirect effects of parasitism, and often observed changes or death rates may be the result of indirect effects of the infection (Poulin, 1995; Timi and Poulin, 2020).

Here, we use LDA to find the combination of features that maximize distinction among the treatment groups of non-parasitized, low, medium, and high parasite load fish. We find that there's a substantial overlap between groups and overall we distinguish non-parasitized fish from low- and medium-parasite load fish mainly by values of respirometry and morphology (FAS, Fineness, SMI), whereas non-parasitized fish and low-parasite load fish are separated from high-parasite load fish mainly by values of kinematics, such as Head:Tail amplitude ratio, Strouhal number, and curvature. More specifically, negative on LD2, high-parasite load individuals show, compared to the population mean, lower tail curvature, and increased H:T amplitude ratio and Strouhal num-

ber (Figure 1.11, and Figure 1.10 for the distribution of the single features); moreover these kinematics are associated with higher MO2 and higher Fineness ratio, and lower U_{crit}. By contrast, orthogonal on the LDA space, low-parasite load fish show lower H:T and Strouhal number (St) than the population mean, in addition to low MO2, high FAS, and high SMI (Figure 1.11 and Figure 1.10 for the distribution of the single features). Although with high variation, low-parasite load fish also include individuals that reach high critical swimming speeds U_{crit}. Similar U_{crit} values are found also for medium-parasite load fish (Figure 1.10).

Fish infected with the ectoparasite *Gyrodactylus* sp. have been previously reported to swim 'erratically' (a "side to side deflections of the body accompanied by sudden body twists referred to as shimmying"), suggesting that the parasites could also affect the kinematics of fish swimming (Cable et al., 2002). Moreover, in later stages of infection, host fins can become contracted and fin rays fuse together, which could indirectly modify the locomotion patterns (Cable et al., 2002; Hockley et al., 2014). Erratic swimming is considered highly visually conspicuous, which suggests it may play an important role in the conspecific recognition of infected individuals and potentially modify shoaling decisions (Bakke et al., 2007; Barber et al., 2000; Krause et al., 1996). In our data, we do not find a uni-directional differentiation of undulatory swimming between non-parasitized and parasitized fish, but using LDA we have an indication of which features better characterize the different treatment groups. For example, among the kinematics, H:T amplitude ratio is an important feature for the first 2 linear discriminants (LDs), together with St and Curvature of the tail. Looking at the H:T amplitude ratio, which can be a marker for the so-called 'erratic swimming', we find that the largest H:T value measured in our data, belongs to the individual with the largest number of parasites (approximately 250 worms) (Figure A.15a, t36 dark red color). For this individual, fins were evidently compromised by the heavy parasite load, which made caudal and pectoral fins stiffer. In addition to this, the same fish also showed an increase in pectoral fin use in steady swimming at higher flow speeds, which is likely to decrease swimming efficiency (Svendsen et al., 2013) (Figure A.15b, t36 dark red). We compare this individual to another parasitized fish (parasite load of 70 worms) that shows a very distinct swimming mode with a very low H:T amplitude ratio at low flow speeds and a constant decrease of pectoral fin use at high flow speeds (Figure A.15, t32 salmon color). This fish also reached one of the highest U_{crit} and counted about 15 worms less compared to the load at the beginning of the experiment, which could also have implications in transmission likelihood (Reynolds et al., 2019). One hypothesis to explain phenotypic variation associated with ectoparasite infection is that parasites might have an effect on the superficial neuromasts, which could in turn modify the ability of the fish to sense the flow and adjust the locomotion

in order to minimize energy expenditure (Montgomery et al., 1997). Taken together, we conclude that ectoparasites, including *Gyrodactylus* sp. can affect swimming behavior and its phenotypic expression, although it's possible that the effect isn't unidirectional, and preconditions or individual variation could determine the variability of the outcome of the infection. In this study, we did not include information on the infection outcome and parasite growth/recovery after the swim tunnel exercise. However, future studies that include more physiological and morphological measures before and after an experiment experiment, could inform us on possible trade-offs between the hosts' response to fight the infection and swimming efficiency.

In addition to characterizing how features vary as a function of infection status, we describe general intra-specific swimming modes within our studied population. Different taxa might use different forms of locomotion, linked to other factors such as habitat, morphology, sex, and more. A recent work compared 46 fish species traditionally divided into classes of anguilliform, sub-carangiform, carangiform, and thunniform swimming modes, but found no distinction on the main kinematic features during steady swimming (Di Santo et al., 2021). This suggests that undulatory swimming is not classifiable into discrete swimming modes, but is rather placed on a continuum (Di Santo et al., 2021). In our study, we use similar morphology and kinematic metrics; to these, we add respirometry measures and found that using PCA we can identify swimming modes described by similar patterns of locomotion and performance. Nonetheless, we also conclude that the modes we describe represent a continuum in the behavioral space and not discrete classes. More specifically, using PCA and hierarchical clustering, we identify three swimming modes that correlate to some extent to swimming performance (*i.e.* vary on the U_{crit} values) (Figure 1.12E, lower right). One main result of the PCA is that we find combinations of morphological, respirometry, and kinematics values that describe differences in the swimming performance (*i.e.* U_{crit} values) (Figure 1.12E, Ucrit). Fish of Cluster 1, which record low U_{crit} values, have a characteristic morphology with high Fineness and low SMI. Fish from Cluster 2 also record low U_{crit} values but their morphological features are similar to the population mean (and to those of Cluster 3). By contrast, fish from Cluster 2 show a larger variation in kinematics, with a high St number, high curvature of the tail, and increased tendency to swim in the back of the swim tunnel (Figure 1.12E). Finally, individuals that reach high U_{crit} values show lower pectoral fin use and St at high flow speeds compared to the mean, lower MO2, and slightly higher body wavelength at higher flow speeds.

One main difference between our study and the study by Di Santo et al. (2021) is that we also include intra-individual locomotion patterns across changing flow speeds, which might hold more information than a single-speed kinematic evaluation (Di Santo

et al., 2021). Another main difference is the analysis on pectoral fin use, which we think affects locomotion patterns (see section 1.3.5 and 1.3.8) (Hale et al., 2006; Svendsen et al., 2013; Webb, 1973). Often, studies on swimming locomotion concentrate on body-caudal fin propulsion, without integrating information on pectoral fin use, which is likely affecting locomotion itself (Hale et al., 2006; Svendsen et al., 2013; Webb, 1973). However, it was previously shown that also for guppies, pectoral fin use can affect swimming locomotion and efficiency. Svendsen et al. (2013) previously found that fish that decreased pectoral fin use early in the trial (at lower swimming speeds), also reduced swimming costs (Svendsen et al., 2013). On the other hand, fish that ceased pectoral fin use at higher swimming speeds, or not at all, reached lower critical swimming speeds (Svendsen et al., 2013). In this study, we also highlight the importance of the evaluation of pectoral fin use as a function of flow speed in the characterization of locomotion patterns. We find that one main difference between a 'bad' from a 'best' performer is also seen in the pectoral fin use: 'bad' performers do not decrease pectoral fin use at the same rate as those individuals that reach high U_{crit} (Figure 1.9, leftmost plot). This result is then confirmed by PCA and clustering (Figure 1.12). Individuals of Cluster 1 measure the lowest U_{crit} values of the population (Figure 1.12E, blue) and record the highest pectoral fin use rate at high flow speeds (Pfin steady in Figure 1.12D, blue); the same fish also record the highest oxygen consumption rates (high MO2 values) (Figure 1.12D, blue). By contrast, individuals of Cluster 2, on average, record low U_{crit} values (Figure 1.12E, orange); however compared to the fish in Cluster 1, fish in Cluster 2 decrease their pectoral fin use rate from low to high flow speeds (Figure 1.12D, orange). Fish of Cluster 2 measure average-to-low oxygen consumption rates (MO2) (Figure 1.12D, orange). Individuals of Cluster 3, on average, record higher U_{crit} values than Cluster 1 and 2 (Figure 1.12E, green) and, at high flow speeds, use pectoral fins at lower rates than the population mean (Figure 1.12D, green). Finally, in our study, we highlight the importance of distinguishing between pectoral fin use at steady versus unsteady swimming states (Figure 1.8). Pectoral fin use serves different purposes in fish's locomotion (Hale et al., 2006); for example, pectoral fins can be used for stabilizing locomotion at low flow speeds, or to contrast loss of balance in particular when swimming at high speeds. To conclude, we confirm the findings of Svendsen et al. (2013), and we also suggest that the pectoral fin modulation across flow speeds, rather than absolute fin use, likely affects locomotion, and consequently swimming costs and performance.

In the context of predictors of swimming performance, we looked at the relationship between morphology and U_{crit}. We find that the correlation between SMI and U_{crit} as well as the correlation between Fineness ratio and U_{crit} is more pronounced for parasitized fish than it is for non-parasitized fish (Figure A.5). This suggests that for para-

sitized fish, morphology might play a more important role in swimming performance for parasitized fish than it does for non-parasitized fish. Swimming endurance varies widely within species (Kern et al., 2018). Despite correction relative to body size, different-sized fish will likely require different levels of activity and thus be hard to compare regardless (Bainbridge, 1958). Fish in our study also vary widely in their swimming performance (Figure A.3) and overall there are limitations to measurements of U_{crit}. For example, it is difficult to compare results across studies due to the different contexts and conditions in which critical swimming speed has been measured in the past. Moreover, it is rare that fish reach their critical swimming speed in a natural context, which also reduces ecological relevance (Plaut, 2001). In future studies, to overcome the limitation of measuring critical swimming speed in artificial environments (such as the flow tank), one should aim at testing things like reproductive success or other behaviors associated with individual measures of endurance or swimming performance. For example, a study by Hammer (1995) showed how the pollutant concentration in the living habitat decreased critical swimming speed (Hammer, 1995). However, in our study we were interested in testing the effect of parasitism on swimming performance, and despite the absence of a clear answer, we see that variation in the population exists, such as for increased U_{crit} in low- and medium-parasite load population (Figure 1.10B), and that it's likely linked to a combination of features that form the multidimensional space of a fish swimming.

In conclusion, in this study, we look at the kinematics that characterize guppy swimming in steady flow conditions and quantify the effect of increasing flow velocity and the role of parasite infection. Among treatments of parasite infections, we don't find strong markers specific to the infection-load category, but we find ranges of variation that are likely associated with each class. These results compare to the findings of Di Santo et al. (2021) in that swimming modes are not discrete categories but are rather found in a continuum in the behavioral space.

Chapter 2

Parasites affect the individual-level behavior of guppies without altering the collective dynamics

2.1 Introduction

Group living animals integrate individual and social information to make rapid decisions in a dynamic environment and the ability to respond appropriately to the behavior of other individuals is essential to minimize exposure to threats and maximize fitness. Parasitism is ubiquitous in animal societies and often hypothesized to be a potential cost of group living (Alexander, 1974), due to the increased risk of disease transmission (Cremer et al., 2007). At the same time, social behavior is also a mechanism by which both vertebrates and invertebrates possibly control, cure, or evade pathogens (Rosengaus et al., 2022). Therefore, the presence of an infectious disease can not only substantially alter the behavior of infected individuals, including the way they interact with others, but also the behavior of uninfected conspecifics. Thus, infection may have cascading effects on individual and collective behaviors. With this chapter, we explore how the infection of a socially-transmitted parasite affects movement dynamics of shoaling guppies, and describe the effects both at the individual and the collective level.

Animals can detect and behaviorally respond to parasitic infections using sensory integration mechanisms both before and after exposure (Lopes, 2022). For example, uninfected individuals can use visual and olfactory cues to detect and limit their exposure to pathogens, such as by avoiding contact with contaminated surfaces or conspecifics (Mikheev et al., 2013; Stockmaier et al., 2021). Avoidance of infected conspecifics occurs in multiple social species (Buck et al., 2018; Croft et al., 2011; Dugatkin et al., 1994; Kavaliers et al., 2004; Stephenson, 2019; Stephenson et al., 2018; Ward et al., 2002), but also

represents a trade-off between the advantages of social living (antipredator defense, foraging efficiency, mating opportunities), and the cost of decreased social cohesion and/or isolation to escape parasites. However, in some cases, it could be difficult to distinguish an avoidance mechanism driven by healthy conspecifics from behavioral changes of the infected host due to the side effects of an infection. Immune-challenged bats, for example, reduce social vocalizations, which consequentially reduce social encounters (Stockmaier et al., 2020); similarly, decreases in social contacts among mice has been found to be a consequence of sickness-induced reduced movement of the infected host (Lopes et al., 2016). Thus, if infection affects the physiological state or the behavioral motivation of an individual, we can expect self-sorting mechanisms to emerge (Couzin and Krause, 2003; Couzin et al., 2002), with the potential to affect social dynamics, independently of direct behaviors initiated by healthy conspecifics.

The aquatic environment has some distinct physical and chemical properties that affect how pathogens are transported and transmitted, and, as a consequence impact responses of both the infected hosts and the uninfected conspecifics (Behringer et al., 2018). Various small fishes including sticklebacks, killifish and guppies have been widely used to study experimental parasitology under lab-controlled conditions (Barber and Dingemanse, 2010; Croft et al., 2011; Krause and Godin, 1996; Lafferty and Morris, 1996; Stephenson, 2019; Sumpter et al., 2008). At the core of these studies, a focus of interest has been developing and understanding the mechanism of the interaction between parasite infection and patterns of behavior. Parasite infection can affect shoaling decisions of infected and non-infected individuals (Barber et al., 1998) and if the parasite is transmitted with contact or proximity, for example, uninfected individuals should avoid the contagious individuals, and choose to shoal with uninfected fish (Krause and Godin, 1996). Previous findings showed that shoals of guppies in which one individual is infected with the ectoparasite *Gyrodactylus* sp. were significantly more likely to initiate shoal fission events by changing their direction of travel (Croft et al., 2011). Similarly, more recent studies on the same host-parasite system also found evidence that guppies can use visual and chemical cues to avoid infected individuals (Stephenson et al., 2018) and that avoidance behavior also depends on the susceptibility of the shoal members, as more susceptible male guppies showed reduced sociality prior to infection (Stephenson, 2019). Given that *Gyrodactylus* sp. are socially transmitted (Richards et al., 2010), such behavior is presumably adaptive as it will be expected to reduce an individual's exposure to the parasites and thus the probability of infection. Here, differently to previous studies, we use cutting-edge tracking technology (Walter and Couzin, 2021) to analyze individual motion features (*e.g.* speed, acceleration, turning rates) of freely shoaling guppies in presence of a socially transmitted parasite, *Gyrodactylus* sp. By maintaining the

individual identity of each fish of the shoal, we quantify the details of the individual movement and interactions, which can vary upon parasitic infection and have cascading effects on the social behavior. This approach overcomes the limitations of binary-choice experimental designs (previously used in Stephenson (2019) with guppies or by Rahn et al. (2015) with sticklebacks), and complements the observational approach previously used for example by Croft et al. (2011) (where fish were observed for a period of 30 minutes).

In this study, we used a novel approach to quantify the group behavior of guppy's fission-fusion movement in presence of a directly-transmitted ectoparasite. To our knowledge, this is one of the first studies using repeated measures of freely shoaling guppies, with mixed infection status, while maintaining the information of the individual identity of each shoal member. The way in which guppies move is highly dynamic and groups split (fission) and merge (fusion) frequently. With our approach, we describe how parasitized and unparasitized individuals behave and interact over time, and we find that parasitic infection increases the likelihood of a set of behaviors to appear. More specifically, infected individuals maintain higher nearest-neighbor distance, spend more time in isolation, are more likely to transition to isolation from bigger group sizes, and tend to occupy more peripheral and frontal positions across different group sizes. Overall, *Gyrodactylus* sp. infection does not seem to induce a single novel disease-specific behavior but rather increases the likelihood for a diseased individual to occupy a space in the extremes of the range of guppy behavioral repertoires.

2.2 Methods

2.2.1 Study system

We used Trinidadian Guppies *Poecilia reticulata*, held in the facilities of the Max Planck Institute of Animal Behavior, Konstanz, Germany. The original stock was received in 2017 by Manfred Schartl from the Univerisity of Würzburg. Holding tanks had circulating water, fish were maintained at 26.5±1 ℃, with average 7.5 pH and 400 μS/cm conductance, under a 12:12 light:darkness cycle and fed daily on live high HUFA *Artemia*, frozen food, including frozen cyclops, frozen *Artemia* and dry food (Naturefood, Supreme Plant - XS and S). All experiments were done in accordance with the permit AZ 35-9185.81/G-17/162 granted by the Regierungspräsidium Freiburg, Baden-Württembeg, Germany.

2.2.2 Infection and screening

For the experimental infections, we initiated a parasite culture of *Gyrodactylus* sp. which were first collected from a local pet shop, and then maintained on guppies from the same experimental fish stock (culture). After each experimental infection, fish were held in a quarantine room, housed individually at a density of 13L per fish in 80L tanks divided into 6 compartments. The compartments were separated by transparent partitions to allow the fish to see conspecifics. Partitions had holes to allow water to flow across compartments.

Wild caught Trinidadian guppies generally host loads of about 10 worms/host of Gyrodactylus spp. (Harris and Lyles, 1992), but can have up to about 270 parasites (Stephenson et al., 2015). *Gyrodactylus* sp. is a monogenean flatworm that reproduces on the skin of the host with a generation time of 24h. Parasite load grows over time but the load on hosts can vary across individuals (Stephenson, 2019; Stephenson et al., 2017), for example with respect to body size, with larger individuals hosting higher parasite loads (Cable and van Oosterhout, 2007; Richards et al., 2010; van Oosterhout et al., 2008).

To initiate infection, first, a culture fish with a heavy infection of worms on the caudal fin was euthanized with an overdose of 0.25 mg/ml MS222 (ethyl 3–aminobenzoic methanesulfonate salt), buffered with sodium bicarbonate. Following death, the caudal fin was cut with a scalpel and moved to a petri dish with fresh water. After about 30 minutes the parasites started detaching from the fin and moved to the water column, either attaching to the bottom of the petri dish or to the water surface (Cable et al., 2002). When enough parasites had detached from the fin, experimental fish were anesthetized with a 0.1 mg/ml dose of MS222, buffered with sodium bicarbonate, and their caudal fin was placed near the floating parasites until five parasites attached to the host's fin. The procedure was observed with a dissecting microscope. After infection experimental fish were revived in fresh water and the number of parasites was counted under a dissecting microscope every 2 or 3 days. In instances where loads were very high but could not perfectly distinguish the precise count, we assigned the value of 200 parasites (for four fish). Fish used for the parasite culture were not used for behavioral experiments.

In order to control for the effects of the manipulation on the fish used for the experimental infection, we collected data by using a sham-infection procedure. Sham-infected fish underwent the same procedure of the experimental infection but no parasites were given to the fish. More details on the control experiment using sham-infection are found in the Appendix B.1, section B.1.1.

2.2.3 Experimental trials

Experimental trials were carried out in 116x116cm custom-built tanks filled with aerated and pH adjusted water and maintained at the temperature of 26.5 ±1 ℃. All the experiments were conducted between October 2019 and July 2020. Videos were recorded for approximately three hours at 30 frames per second using a high-resolution camera (Ximea MC124MG-SY-UB) with a 16mm lens (Kowa LM16HC 16-mm f/1.4). Data acquisition was controlled by video recording software (Motif, loopbio GmbH). The illumination system was custom-built as an array of 15x15 infrared LEDs (Winger IR power LED 850nm 3W). Figure 2.1 shows four of the setups used for data collection.

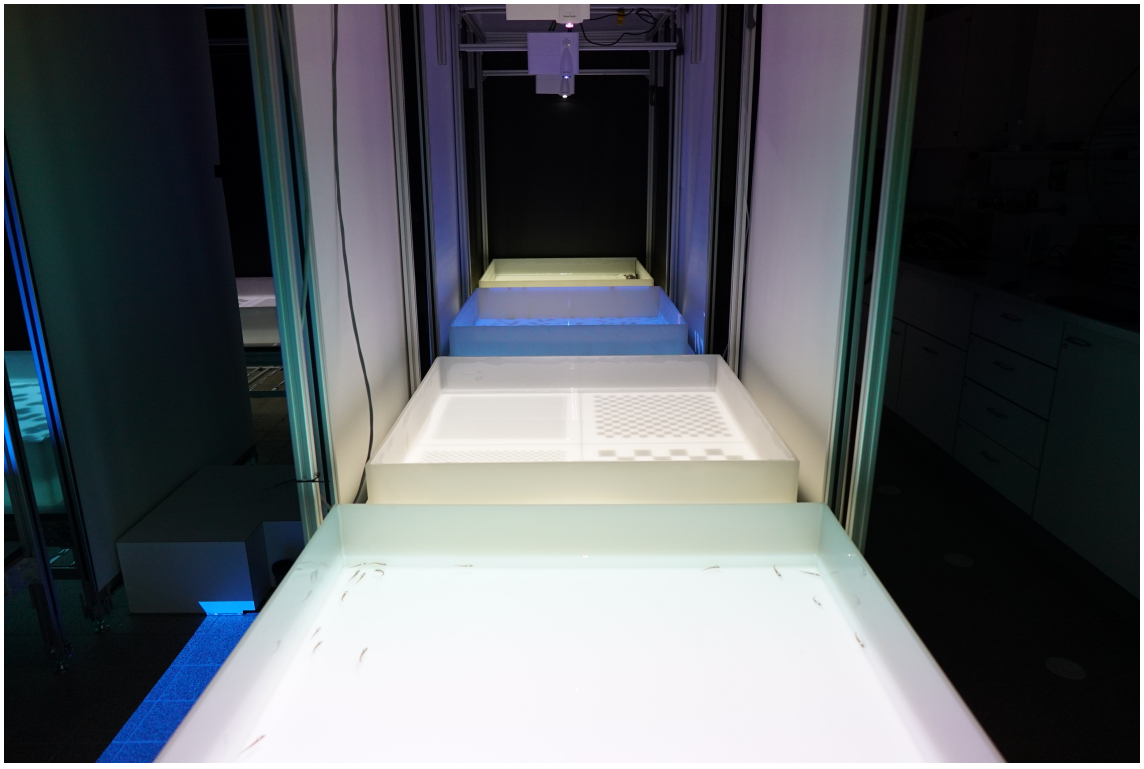


Figure 2.1: **Experimental arenas.** Picture of the experimental arenas: 116x116cm custom-built tanks filled with aerated and pH adjusted water and maintained at the temperature of 26.5 ±1 ℃. Cameras were placed above the tank and videos were recorded at 30 frames per second. Data acquisition was controlled by video recording software (Motif, loopbio GmbH). The illumination system was custom-built as an array of 15x15 infrared LEDs (Winger® IR power LED 850nm 3W) and was placed below the tanks.

An experiment consisted of a Baseline (BSL) and a Treatment (TRT) condition, in order to record the same fish (the Focal-fish, FF) at two time points: before parasite infection (Baseline) and after parasite infection (Treatment). In both conditions, the Focal-fish was recorded for about 3 hours while shoaling with seven other non-parasitized fish, that is the Non-Focal fish (NF) (Figure 2.2A, B). Throughout the chapter, one condition (Baseline or Treatment) counts as one trial. During the Baseline condition, all fish were

non-parasitized. The Focal-fish was recognized by an elastomer tag which allowed visual identification at the end of each recording. At the end of the Baseline condition, the Focal-fish was anesthetized and underwent a procedure to initiate parasite infection (Figure 2.2A Infection procedure; for details on the infection procedure see section 2.2.2). After infection, fish were kept in semi-isolation until the following testing day, that is the Treatment condition (Figure 2.2A to B). Semi-isolation was used to allow the parasite infection to grow undisturbed by physical contact of conspecifics in proximity, while maintaining visual contact. The Treatment condition was collected after either 5, 10, or 15 days, which corresponded to a either short (5 days), medium (10 days), or long (15 days) infection duration, which in this study on average corresponded to low, high and reduced (after a peak) infection loads (Figure 2.3). Besides the absolute number of worms per fish, for each individual and the parasite growth curve, we calculated two indices, the growth rate and infection integral. The growth rate indicates how infection grew from the first day of the infection and it is calculated as ln(parasites day 1) - ln(parasites day experiment)/n day experiment. The infection integral is instead calculated as the area under the curve from the first day of infection until the day of the experiment. Each fish was randomly assigned to either one or two Treatment condition days and was recorded for a maximum of 3 times, including Baseline and 1 or 2 post-infection time points.

The Treatment condition consisted of the same recording procedure as the Baseline, that is 3 hours of consecutive recording of one Focal-fish, in this case parasitized, and 7 Non-focal fish (Figure 2.2B). At the end of each Treatment condition trial, each fish was removed from the tank one-by-one and placed in individual casing in a box with transparent dividers (27x15cm, with single partitions inside for each fish). Individual housing was used to avoid potential parasite transmission by forcing proximity when transporting fish from the experimental area to the quarantine holding area. All eight fish were then anesthetized with a 0.1 mg/ml dose of MS222, buffered with sodium bicarbonate, and screened for parasites immediately to asses possible infection transmission (Figure 2.2B, screening for parasites). After screening, fish were returned to fresh water and let recover from the anesthetized state, which usually took a few seconds for full recovery. If at least one individual was infected during the trial, the eight fish were filmed for a further 15 minutes, while keeping them separate in the box (Figure 2.2B, bottom right). The fish inside the box with dividers, were filmed using the same camera hardware and lighting. This procedure allowed post-experiment individual recognition, thus the ability to determine the precise identity of all the individuals that were infected during the trial. The post-experiment identification is achieved with a method known as 'transfer learning'. This is a machine learning technique that uses knowledge learned from previous neural network training (in our case from the 3 hours recording of the same fish

shoaling during the trial) and applies it to a new situation (in this case the recording of the same fish, but in the box and after the experimental recording was stopped for some minutes). The same experimental procedure was used for control experiments, where the Focal-fish was sham-infected (Appendix B.1).

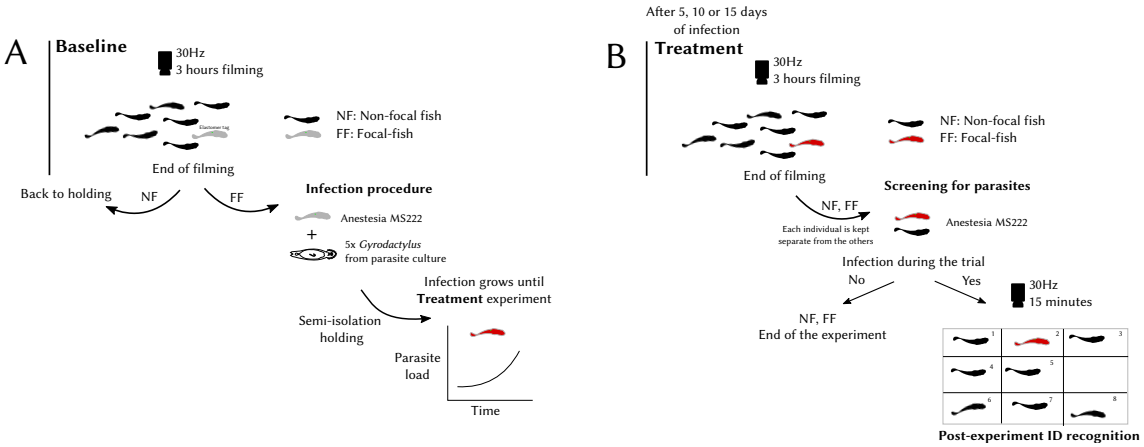


Figure 2.2: **Illustration of the experiment design.** Each experiment consisted of a Baseline (A) and a Treatment (B) condition, and in each condition, we recorded eight guppies freely shoaling, divided between seven Non-focal fish (NF) and one Focal-fish (FF), for 3 hours at 30Hz. In the Baseline condition, all fish were non-parasitized, and the Focal-fish was recognized by an elastomer tag. At the end of the Baseline, the NF went back to holding, while the FF underwent the infection procedure. The FF was anesthetized was infected with five parasites (see more about the infection procedure on section 2.2.2). At the end of the infection procedure, the FF was revived in fresh water and housed individually; each FF was screened every 2-3 days to note the parasite count. Each FF was then filmed a second time in the Treatment (B) condition, with seven NF, randomly selected (and therefore unlikely to be the same as for the Baseline condition). At the end of the Treatment recording, each fish was removed one-by-one, placed in a box with dividers for each individual, anesthetized, and screened for infection that might have occurred during the trial. If infection happened, the fish were filmed for 15 minutes, while still separated in the box (B, bottom right), to allow post-experiment identity recognition.

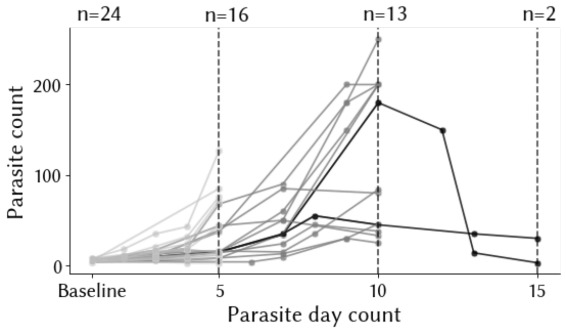


Figure 2.3: **Parasite load growth over time.** Parasite load and its growth from Baseline to the testing day in the Treatment condition. Each line represents one fish measured on different days. The vertical line represents the different possible Treatment days. In this experiment, the growth shows approximately exponential growth within the first 10-15 days of a naive infection, followed by a decay over the following days. In our trials, the two individuals tested on day 15 both had lower counts on day 15 of testing. More details on the experimental design and the day of testing are found in section 2.2.3.

We recorded a total of 76 trials, of which we fully processed 55 trials. A total of 21 trials were excluded from further processing in case of low tracking quality or absence of the Treatment condition upon death or poor body condition (due to the infection) of the infected fish prior to the Treatment testing day. In one case, a fish deceased during the recording, following which the trial was interrupted and the data were not processed. We included 55 divided between Baseline (BSL, N=24) and Treatment (TRT, N=31) conditions. From the Treatment condition, we counted N=16 at day 5, N=13 at day 10, N=2 at day 15. A total of 6 fish were recorded at two experimental time points (that is both at day 5 and day 10, or day 10 and day 15). In each trial, we used eight size-matched females. We restricted the analysis to female guppies as they form natural schools more so than males, and we wanted to exclude the emergence of secondary effects induced by mating behavior (Croft et al., 2011; Griffiths and Magurran, 1998; Stephenson, 2019). Although in the wild the density of guppies varies depending on predation pressure, season, location and other factors (Magurran, 2005), guppies on average form shoals of size between 4 and 40 individuals (Croft et al., 2003; T. Russell et al., 2004); therefore our tested group size is representative of the natural population density of guppies in the wild (Croft et al., 2011; Magurran, 2005).

2.2.4 Tracking and data processing

We tracked all videos using TRex (Walter and Couzin, 2021), that allows individual recognition and identification for unmarked individuals (Figure 2.4). For each recording, we exported raw x, y centroid coordinates of all individuals. Video analysis of the main experimental trial was restricted to the last two hours of the experiment, as the first

hour of the experiment includes tracking artifacts due to general acclimation to the tank environment. On average our guppy shoals seemed to reach a consistent swimming speed, nearest-neighbor distance, and group size preference after the first hour (Figure 2.5). The videos used for identity recognition after infection were analyzed separately, and the identity of fish that got infected during the trial was noted and included for the following analyses.

Note that raw x, y coordinates exported from the tracking program may be missing values for some frames for a multitude of reasons, including fish crossing/overlapping, fish near the wall, sudden identity switches, and general centroid-tracking noise. However, we found that in the instances after identity was momentarily lost or switched, when tracking was resumed, fish were correctly identified by the neural network that TRex uses for maintaining individual identities (Walter and Couzin, 2021). In the time tracking is lost before re-identification, however, TRex by default deliberately assigns invalid values for an individual's x, y coordinates. To simplify the process of data analysis, we filled in these missing values using linear interpolation, and the interpolated values were used to inform further filtering in a secondary step (see section 2.2.5).

Before filtering and after interpolating, we calculated individual behavioral motion metrics (speed, acceleration, heading), and pairwise distances for all individuals (see next section 2.2.5). To partly account for inevitable (albeit minor) tracking noise, we applied a step in the derivative for the calculation of the behavioral metrics (speed: 4 frames, acceleration: 2 frames, heading: 2 frames). We calculated speed and acceleration assuming a constant frame rate, but this was not always the case due to hardware problems (on average 0.07 %, Figure B.3). Therefore, as a first filtering step, we identified frames where inter-frame interval was different from 1/frame rate, and assigned invalid numbers to speed and acceleration for the detected 'skipped' frames. As a second filtering step, we used speed values to identify tracking errors. This is efficient when fish are assigned the wrong identity, but are distant from each other. In these instances, the calculated speed value is abnormally high, which we therefore detect in order to avoid further errors based on the evaluation of metrics on frames where the identity of the individuals is uncertain. To this end, we identified clear 'jumps' in the x, y coordinates, detected by high speed values, that could be caused by temporary identity matching errors (here instances of swimming speed higher than 35cm/s with a minimum duration of 1/10 of a second). When two or more such instances were found to be separated by less than 5 frames (1/16 of a second), they were combined into one. Each of the identified windows was padded with 8 frames to account for the beginning and the end of the possible tracking error. Frames identified with the tracking-error filter were excluded from further calculations. Note that in case of mixing of identities of individuals in close

proximity, we have a low chance of correctly identifying those tracking errors. However, due to the overall high reliability of the tracking, such small tracking errors very unlikely impact our analyses.

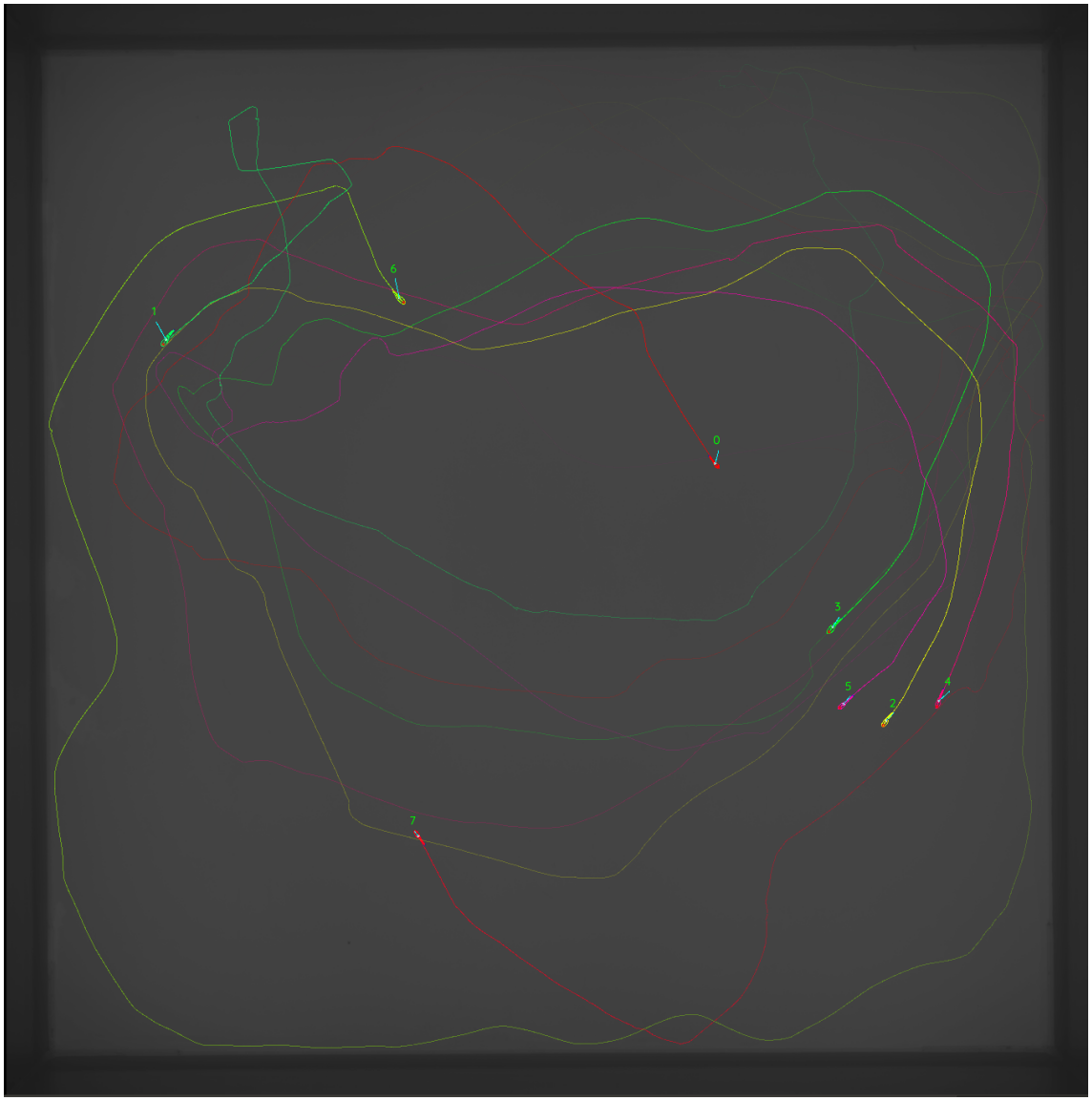
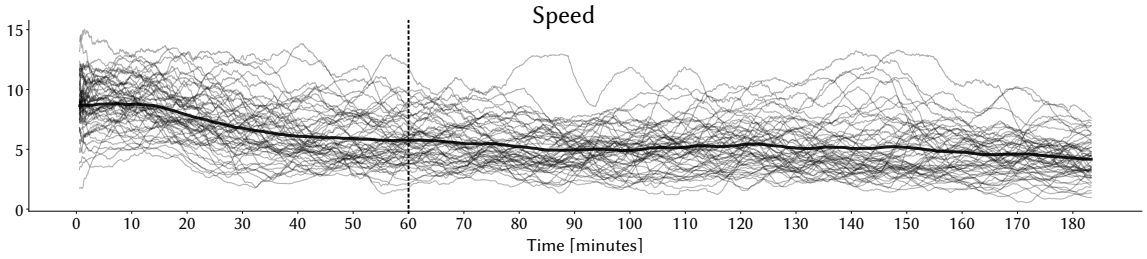


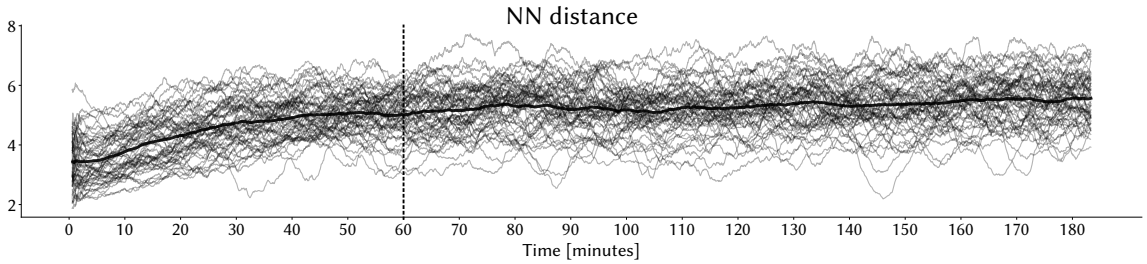
Figure 2.4: **Camera view of the arena with tracked individuals.** Picture of the experimental arena with individual fish tracked with TRex, where each color corresponds to one fish, which identity is maintained throughout the recording (Walter and Couzin, 2021).

2.2.5 Pairwise distance, group sizes, events

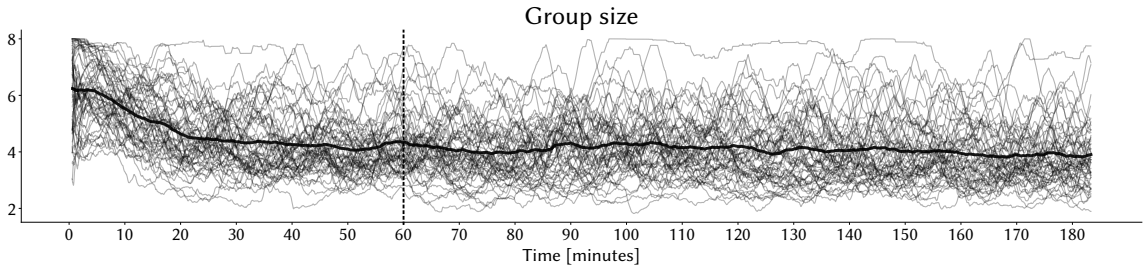
Guppies, like many other social species, live in fission-fusion populations; thus their collective behavior is highly dynamic, characterized by continuous splitting (fission) and merging (fusion) of groups (Couzin and Laidre, 2009; Wilson et al., 2014). Therefore, we developed a workflow to identify groups of fish (characterized by group members swim-



(a) Speed of all fish and of all trials over time (moving average of 5 minutes)



(b) Nearest-neighbor distance of all fish and trials over time (moving average of 5 minutes)



(c) Group size of all fish and of all trials over time (moving average of 5 minutes)

Figure 2.5: **Stabilization of behavioral metrics over time.** (a) Speed (cm/s), (b) nearest-neighbor (NN) distance (cm), and (c) group size (of the event) calculated with a moving average with a window size of 5 minutes. Each line in each figure represents values for each fish in each trial. The thinker black line represents the overall average across all fish and trials. The metrics displayed show a visible acclimation effect over the first hour of recordings. The dotted line marks the first hour of the experiments; data analysis was restricted to the last two hours of the experiment due to a visible change in the variables in the first 60 minutes of a trial, that is the increase of nearest-neighbor distance (b) and the decrease of speed (a) and average group size (c).

ming in close proximity to one another), evaluate the number of fish in each group, and define the timing corresponding to the beginning and the end of each group, detailed as follows (Figure 2.6). The first step was to calculate pairwise distance across all fish at each time point, used to identify distance-based aggregations and assign group membership (Figure 2.6A). A 'group' is an aggregation of fish that can co-occur with other spatially separated aggregations, and group size can be between 1 and 8 individuals, where 8 is the maximum number of fish in an experiment, and 1 corresponds to an isolated indi-

vidual (Figure 2.6B). For the identification of one or more groups in the same frame, we used pairwise distance to define a 'network' of connected individuals and specified a group as connected components within this network. In this procedure, one individual is connected to another if they are located within a distance threshold. To define a relevant distance threshold for our data, we calculated the 5^{th} percentile of the median pairwise distance across all data and used this value, which was 15cm. Figure 2.6B shows a simplified version of the group identification process where the circle around the fish in the figure represents the distance threshold of 15cm. The number of fish in each group determines what we refer to as 'group size' (Figure 2.6C). The assignment of group size depends on the threshold we used for defining groups, as a bigger threshold will allow groups to be large and last longer. Because small fluctuations of distance near the threshold can lead to noise in the group membership assignments, we used a moving average on the distance values with a window size of 1 second. The last step is to identify what we call an 'event'. An event is defined as a subset of coordinates of consecutive frames where one or more individuals are part of the same group (Figure 2.6D). To define an event, we combine trajectories and identity information based on the group classification (Figure 2.6B, C). An event ends when the group composition changes, that is when one (or more) individuals leave the group (fission), or when one or more individuals join the group (fusion), thus changing the group size and/or composition. Each event is therefore characterized by a specific number of individuals (*i.e.* group size), composition (based on the identity of the individuals), and duration.

Based on the event detection, we excluded some events based on further filtering thresholds. We excluded events containing more than 10% of the frames with possible tracking errors and/or interpolated values of the raw trajectories (as explained in section 2.2.4). This method is rather conservative and excludes, in one extreme case, up to 25% of the data (in one case of a trial with poor tracking quality, with individuals not moving much). However, the median percentage of data excluded due to tracking errors is 0.6% of the frames (Figure B.4). Overall, multiple processing steps are necessary to simplify the analysis and to create robust measures of behavioral metrics that minimize tracking errors. These include smoothing pairwise distance calculation, exclusion of events with tracking errors, as well as excluding short events.

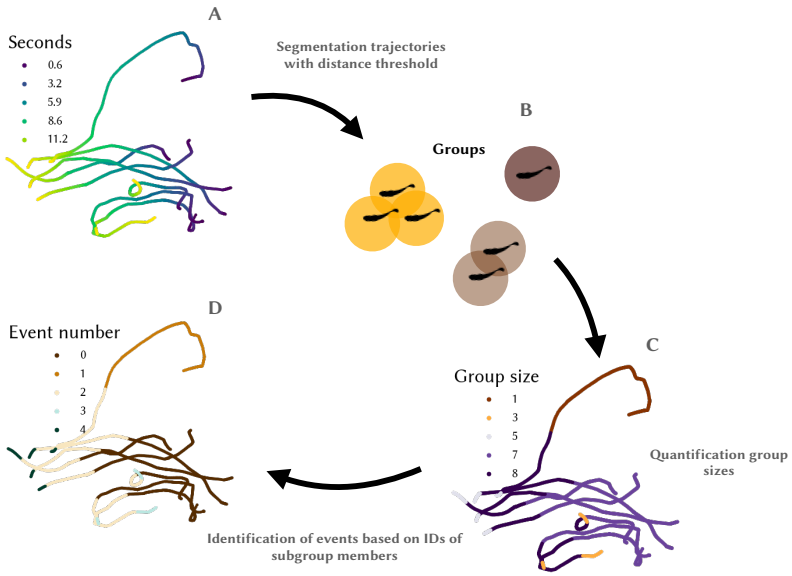


Figure 2.6: **Work-flow of data processing from trajectories to events segmentation.** Work-flow of segmentation of trajectories of freely swimming individuals of trajectories into one or more groups and events over time. (A) Trajectories of a subset of eight individuals over a time of 12 seconds. (B) Groups are identified based on a pairwise distance threshold of 15cm (for more details, see section 2.2.5). (C) Trajectories are segmented into one or more groups of different sizes; color denotes size. (D) Segmented trajectories are divided into events, where both group size and composition are taken into consideration.

2.2.6 Behavioral metrics

We calculated behavioral metrics for each individual within each event and trial. More specifically we calculated individual-based metrics of speed, acceleration, heading angle, nearest-neighbor distance, speed relative to the group mean, nearest-neighbor alignment, distance from the tank wall, and group-level metrics, such as group centroid speed, acceleration, and polarization. All the metrics are then calculated as averages and medians across trials and across fish in the trial, in order to obtain one value per-fish, per-trial. In the analysis, we only included frames where all eight individuals are correctly tracked and identified, in order to avoid erroneous interpretations associated with relative measures and estimation of group sizes.

2.2.7 Fission-fusion analysis

After the segmentation of trajectories into events, we defined metrics to quantify fission-fusion. In particular, we determined properties of the group sizes that dynamically vary throughout the whole trial. First, we defined the group size distribution at each time point. The group size distribution at time t is represented with G_{tk}, where k is an index that goes from 1 to the N_t, which is the number of groups at time t. The largest group

has index of $k = 1$, and the smallest group has index $k = N_t$. To illustrate this notation, consider a time t when all fish are in a single group: then $N_t = 1$ and $G_{t1} = 8$. At a time t when there are three groups of sizes 4, 3, and 1, we have $G_{t1} = 4$, $G_{t2} = 3$, and $G_{t3} = 1$, and $N_t = 3$.

In order to determine the ending of an event, we look at the changes in the group membership in each time instance. The membership of each group is defined with M_{tki}, where $i = 1...G_{tk}$. A change in group configurations is noted when $N_{t+1} \neq N_t$, or more generally when $M_{t+1,ki} \neq M_{tki}$. The proportion of time one fish spends in the different group sizes is calculated as the sum of the frames in the group G_n divided by the total number of frames in each of the 8 possible group sizes. In order to examine how individuals differ from other fish in the same trial, the value of the proportion of time each fish spent in each group size is normalized by subtracting and dividing by the trial mean. The normalization is to account for differences between trials in the amount of time fish spend in the different group sizes.

Furthermore, for each fish, we quantified the likelihood of the change from one group size to the next, here referred to as a 'transition'. This is a way to quantify the the average group size an individual transitions to, after a fission or a fusion (that is after an event ends), and the associated probability. In this analysis, we excluded all transitions that included invalid values (determined by previous filtering on tracking errors, see section 2.2.4).

2.2.8 Spatial positioning

Based on the position and direction of motion of the centroid of each group in each frame, we calculated the relative position of each individual with respect to the front-back and the center-periphery axis. For the front-back position calculation, we rotated each individual orientation with respect to the group centroid movement direction (Couzin et al., 2002; Katz et al., 2011). The center-periphery distance is simply inferred based on the distance from the group centroid. The analysis of the position of each individual in the group is restricted to events with a minimum group polarization value of 0.65.

For each individual, we calculated a categorical position value by ranking the front-back and the center-periphery position, where small values indicate the front and the center of the group, respectively. For example, for a group size of 5, in the front-back ranking a value of 1 indicates the front-most individual while 5 indicates the rearmost, and in the center-periphery ranking a 1 indicates the fish closest to the group centroid and 5 the fish farthest from the centroid. Finally, for each individual, we calculated the proportion of time they spent in each relative front-back or center-periphery position, and more specifically some analysis was restricted to the proportion of time each indi-

vidual spent in the most front and the most peripheral location (as determined by the group size).

In addition to the relative position in each group size, for each individual, we also calculated the nearest-neighbor probability distribution. For this analysis, we generated a nearest-neighbor adjacency matrix A_{ij} to represent the nearest-neighbor associations at each frame based on the event's group members. The nearest neighbor is marked with a 1, while all the others are marked with zero (Figure 2.9a. When an individual is in a group size of 1, the diagonal value is set to 1 (as in to be the nearest-neighbor to itself). To estimate the probability we calculated the mean of the A_{ij} matrix over the whole duration of the experiment excluding the diagonal.

2.2.9 PCA

We used principal component analysis (PCA) to extract the dominant axes of variation. The behavioral metrics used are: relative speed, nearest-neighbor distance, nearest-neighbor alignment, proportion of time spent in the front, proportion of time spent in the periphery, probability of transition to a group size of 1, and proportion of time spent in the different groups sizes. With these metrics, we create a data matrix M_{ij}, where each row i represents one fish in each group size (between 2 and 8 for all the metrics except for the proportion of time in the different group sizes, restricted between 1 and 7), and columns j = 1..8 are the different quantities. Following standard procedures, we normalized the data matrix **M** so that the column mean is zero and the column standard deviation is 1. We then performed principal component analysis (PCA) on the resulting matrix.

2.2.10 Data analysis and statistics

Data were analyzed with custom-written code in Python. For the logistic regression models, we use the package `statsmodels` and the `statsmodels.logit` function. To evaluate the results of the logistic regression fitting, we used Average Marginal Effects (AMEs) (Leeper, 2017). Correlation coefficients were calculated using the Pearson Correlation Coefficient using the function `scipy.stats.pearsonr`. In the results section, all correlations are calculated as Pearson correlation coefficient, reported with the significance p-value. Other statistical tests are reported in the text in the results section.

2.3 Results

The results are organized as follows:

1. Parasite load and fish size (Section 2.3.1)
2. Trial-level differences: How do Baseline trials differ from Treatment trials (Section 2.3.2)
3. Individual differences in a fission-fusion system (Section 2.3.3)
4. Speed and group size distributions (Section 2.3.4)
5. Dynamics of group size changes (Section 2.3.5)
6. Physiological and behavioral predictors of parasite transmission (Section 2.3.6)
7. PCA on the Focal-fish behavioral metrics (Section 2.3.7)

2.3.1 Parasite load and fish size

Figure 2.3 shows the curve of the growth of the parasites on all tested individuals. The dotted vertical lines mark days in which Focal-fish from the Treatment condition were tested. In our dataset, parasite load seems to reach the highest values around the 10^{th} day of infection. However, we only tested two fish on the 15^{th} day of the infection, and in both cases, we recorded lower infection loads than the previous testing days.

Figure 2.7 (top row) shows how body size, parasite growth rate, infection integral, and parasite load distribute with respect to infection duration. For an explanation regarding how growth rate and infection integral are calculated, see section 2.2.3. Although the trial procedure placed fish randomly in size-matched groups, we found a small bias in Focal-fish having a relatively larger body size compared to the trial mean for trials held on day 10 and 15 (Figure 2.7, left). Parasite growth rate is similar at both days 5 and 10 of the experiment, with the growth rate decreasing at day 15. This is, however, based on data from only two individuals, and in our case, this coincides with a decrease in the infection load. Because it integrates over time, infection integral increases with increasing infection duration, and parasite load is generally largest on day 10 of the experiment, where we also record the largest variance. We do not find correlations between body length and parasite growth values.

2.3.2 Trial-level differences

To obtain an overview of trial-level effects of the presence of a parasitized individual in the shoal, we first compared group metrics between all Baseline and Treatment trials (Figure 2.2). In total, we compared eight metrics (Figure 2.8). First, to quantify how the group size distribution varied over the course of a trial, we calculated the average

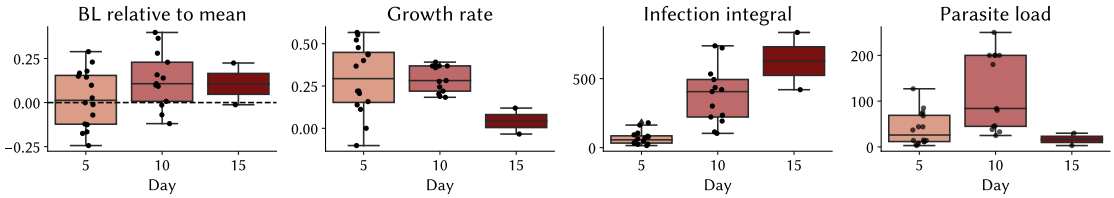


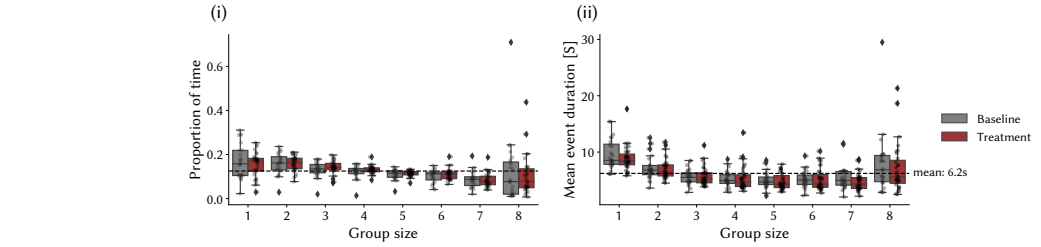
Figure 2.7: **Morphology and parasite-growth quantities as a function of time of the experiment.** Body length of the Focal-fish relative to trial mean, parasite growth rate, infection integral, and parasite load. Each plot shows the distribution of each metric as a function of testing day. Box plots give the median, interquartile range (box), and values that sit outside of 1.5 times the interquartile range (whiskers).

proportion of time a fish in each trial spent in the different group sizes, as well as the average time in seconds spent in each group size (Figure 2.8a). In this comparison, we find no difference across conditions (Baseline to Treatment), which suggests that the presence of a single parasitized individual does not alter general spatio-temporal dynamics. On average, across all trials, fish spent more time in smaller groups compared to larger groups; more specifically, guppies spent 16% of the time in a group size of one, thus isolation, compared to the time spent in a group that included all fish in a trial (group size of 8), that is 11% of the whole trial duration (Figure 2.8a i) ($t=2.702$, $p<0.01$). On average, a unique group configuration (or event, see more in the Methods section 2.2.7) lasted 6.2 seconds (s.d. 2.8), averaged across group sizes (Figure 2.8a ii). The group size of 1 shows a longer duration than the average of the other group sizes (8.5s ±2.4, Figure 2.8a ii).

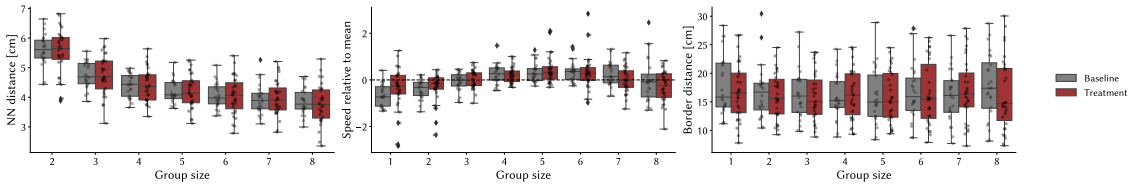
We also quantified metrics of movement, neighbor alignment, and polarization, averaged across all fish in a trial. More specifically we compared median nearest-neighbor distance, speed relative to the trial mean, median distance from the wall, speed median, median group polarization, and median alignment to the nearest-neighbor polarization. Overall, looking at the averages for each trial, we do not find major differences in the distributions of the metrics of median nearest-neighbor distance, speed relative to the trial mean, median distance from the wall, speed median, median group polarization, and median alignment to the nearest-neighbor polarization between the Baseline and the Treatment condition (Figure 2.8b, and Figure 2.8c).

2.3.3 Individual differences in fission-fusion

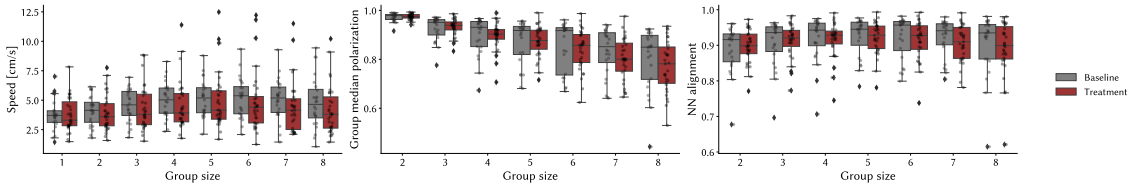
In the previous section we looked at metrics including all groups members - now we focus on individuals, and specifically compare the Focal-fish to the Non-focal fish at the Baseline and at the Treatment condition. Therefore, unless stated differently, the following results compare values on metrics calculated for each individual, calculated



(a) Temporal properties of group sizes across trials



(b) Nearest-neighbor distance, relative speed, border distance



(c) Speed, group polarization, nearest-neighbor alignment

Figure 2.8: **Trial-level comparison.** Distribution of (a) the temporal properties of group sizes across trials and (b,c) behavioral metrics and variation with respect to group size. Each box plot shows the distribution averaged across all individuals in a trial, each dot represents one trial. (a, i): Distribution of appearance likelihood for each group size. Dotted line denotes a uniform probability of appearance with 1/8 likelihood. Most group sizes are represented at similar rates, although overall guppies spent more time in a group size of one (16% of the time), compared to the time spent in a group size of 8 (11% of the time). (a, ii): Mean event duration in seconds, averaged across trials. The group size of one has a higher mean across group sizes (8.5 s ±2.4). The dotted line marks the overall mean across group sizes, that is 6.2 seconds. (b, c) Distribution of behavioral metrics and variation with respect to group size. (b) From left to right: nearest-neighbor distance, speed relative to trial mean, distance from the wall. (c) From left to right: speed, group polarization, nearest-neighbor (NN) polarization. Overall, we do not find major differences in the distributions of the metrics (b,c). Box plots give the median, interquartile range (box), and values that sit outside of 1.5 times the interquartile range (whiskers).

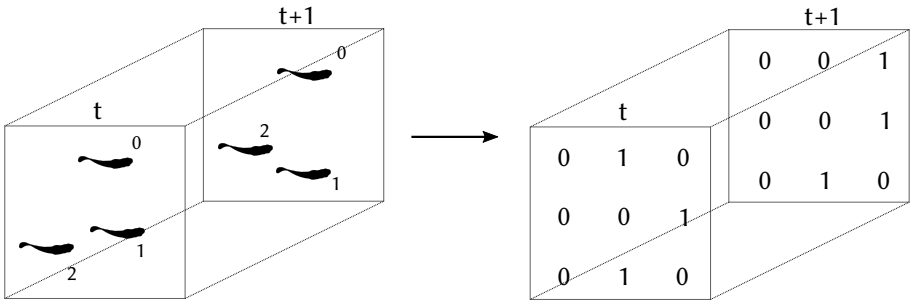
for each group size the fish is in.

Nearest-neighbor configuration

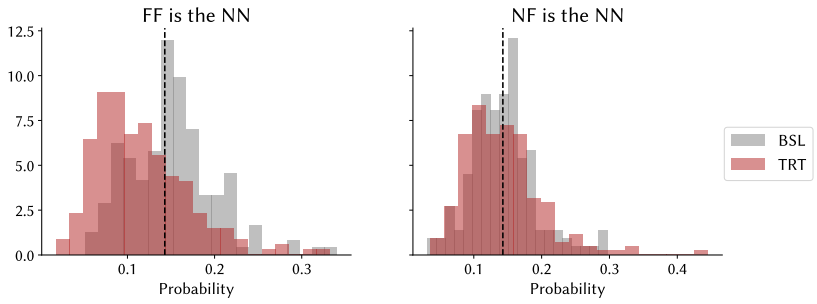
In order to understand how the presence of infected individuals affects group composition and positioning, we first looked at the relative position of individuals within a group. More specifically we asked whether the disease state of the Focal-fish affects its position with respect to others. This analysis includes two parts: 1) likelihood of each fish to be the nearest neighbor of any other fish; 2) relative position within the group

in terms of front-back and center-periphery positioning. For this analysis, we used only frames that are common to all fish.

The expected value for unbiased group position sorting, where each individual is equally likely to be the nearest-neighbor, corresponds to $1/7$, that is 0.14, where 7 is the number of possible neighbors. We compared the probability for the Focal-fish to be the nearest-neighbor of any other fish to the same probability for the Non-focal fish. We find that in the Treatment (TRT) condition, the Focal-fish is significantly less likely to be the nearest-neighbor of Non-focal fish (mean = 0.1, s.d. = 0.05), compared to the Baseline (BSL) condition (mean = 0.15, s.d. = 0.05) (Kruskal-Wallis H-test, test statistic=53.558, $p < 0.001$). On the other hand, we looked at the probability for the Non-focal fish to be the nearest-neighbor of the infected Focal-fish. In this case, we do not find a difference between the BSL (mean = 0.14, s.d. = 0.05) and the TRT (mean = 0.14, s.d. = 0.06) condition for the Non-focal fish (Figure 2.9).



(a) Explanation of calculation of the nearest-neighbor associations



(b) Probability to be the nearest-neighbor

Figure 2.9: **Group composition: nearest neighbor distribution**. (a) How we built the nearest-neighbor adjacency matrix. In the matrix, the nearest neighbor is marked with a 1, and anything else is set to 0. (b) Left: Distributions of the probabilities for the Focal-fish to be the nearest-neighbor of any other Non-focal fish, comparing Baseline (gray) to Treatment (red). Right: distributions of the probabilities for any Non-focal fish to be the nearest-neighbor of the Focal-fish, comparing Baseline (gray) to Treatment (red). Dotted line represents the probability of 1/7, that is the probability of fish having no preferred nearest neighbor. BSL: Baseline; TRT: Treatment

Ranked spatial position in the group and relationship to isolation

The spatial position of each individual in a group was calculated with respect to two axes: 1) center-periphery, looking at the probability associated with the radial distance from the group centroid; 2) front-back, calculating the probability for each fish to occupy either frontal or posterior positions. Each center-periphery and front-back position was ranked with respect to other individuals in the group. To calculate the center-periphery distribution, we ranked the distance from the centroid at each point in time. Similarly, we ranked the front-back position distribution by ranking the position relative to the orthogonal axis of the group centroid.

Spatial positioning in a group has been previously used as a measure of local-interaction rules that depend on the motion of the neighbors and determine spatial configuration (Couzin et al., 2002; Katz et al., 2011). To visualize preference differences in group position, we calculated the proportion of time each fish spends in each rank position, with respect to the group size and the Treatment (Figure 2.10). Figure 2.10a shows the probability associated with center-periphery rank positions (low values: center; high values: periphery), while Figure 2.10b shows the values for ranking in front-back positions (low values: front; high values: back). For each ranking calculation, we compare the Baseline values (panels i and ii of Figures 2.10a and 2.10b) to the Treatment (panels iii and iv of Figures 2.10a and 2.10b); the columns of the figure separate Non-focal fish (left, panels i and iii) to Focal-fish (right, panels ii and iv). Each matrix shows values of the normalized proportion of time at each ranking (x-axis) for each group size (y-axis); therefore the largest ranking value always corresponds to the group size. For a visualization of the variance associated with the proportion of time spent in each center-periphery ranking position, see Figure B.5a and Figure B.5b.

Figure 2.10a shows that in the Baseline condition, we see little to no difference between the Non-focal fish (Figure 2.10a i) and the Focal-fish (Figure 2.10a ii). In the Treatment condition we see that the Non-focal fish spatial positioning remains unchanged with respect to the Baseline (Figure 2.10a iii), whereas the Focal-fish shows a marked increase in the likelihood of being in peripheral ranking positions of the group, for all group sizes (Figure 2.10a iv). Figure 2.10b shows that in the Baseline condition, we do not see a bias for the Non-focal fish with respect to front-back ranking of positions across group sizes; Focal-fish in the Baseline conditions show a slightly reduced probability to occupying frontal positions compared to other ranking positions (Figure 2.10b i). In the Treatment condition, while the Non-focal fish show no preference for any front-back ranking, the Focal-fish shows a greater likelihood to be in frontal positions (Figure 2.10b, iv). For a visualization of the variance associated with the proportion of time spent in each front-back ranking position, see Figure B.5c and Figure B.5d.

The relative position of the focal individual in the group correlates with other behavioral metrics, including the time fish spend in isolation. Individuals that are more peripheral in large groups, spend more time in isolation (Figure B.6). However, we find little to no correlation between time spent in frontal ranking positions and the time spent in the group size of one (Figure B.7). Together, this indicates that individuals at the back of a big group, or distant from the group centroid, have an increased likelihood to also spend more time in isolation than to individuals that on average spend more time in central positions of the group.

Nearest-neighbor distance and nearest-neighbor alignment

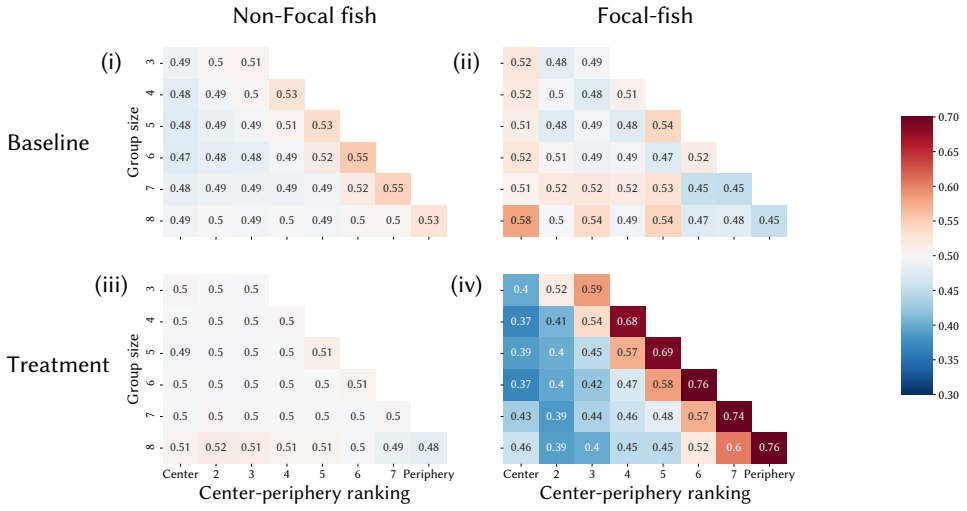
In this section, we compare the average nearest-neighbor distance and the nearest-neighbor alignment between the Non-focal fish and the Focal-fish, in the conditions of Baseline and Treatment. The values of nearest-neighbor distance and the nearest-neighbor alignment are calculated for each individual separately. Nearest neighbor distance is an indicator of the minimum distance individuals keep with their nearest neighbors. We find little to no difference in the nearest-neighbor distance distribution across group sizes between Focal- and Non-focal fish in the Baseline condition (Figure 2.11a, Baseline). On the other hand, we find that parasitized Focal-fish measure higher nearest-neighbor distance, across group sizes, in the Treatment condition, compared to the Non-focal fish (Figure 2.11a, Treatment). In order to verify that changes in nearest-neighbor distance were not driven by effects of the manipulation on the fish in the experimental procedure, we compared these results to a control experiment (Appendix B.1), where the Focal-fish was sham-infected (Appendix B.1.1). We found that the nearest-neighbor distance did not increase as a consequence of the sham-infection manipulation on the Focal-fish in the treatment condition (Figure B.1).

As a measure of the alignment among pairs of individuals, we looked at values of nearest-neighbor alignment (Figure 2.11b). The mean nearest-neighbor alignment is higher in the Baseline condition compared to the Treatment, although in both cases the confidence intervals overlap and therefore the difference between Focal and Non-focal fish is not significant (Figure 2.11b).

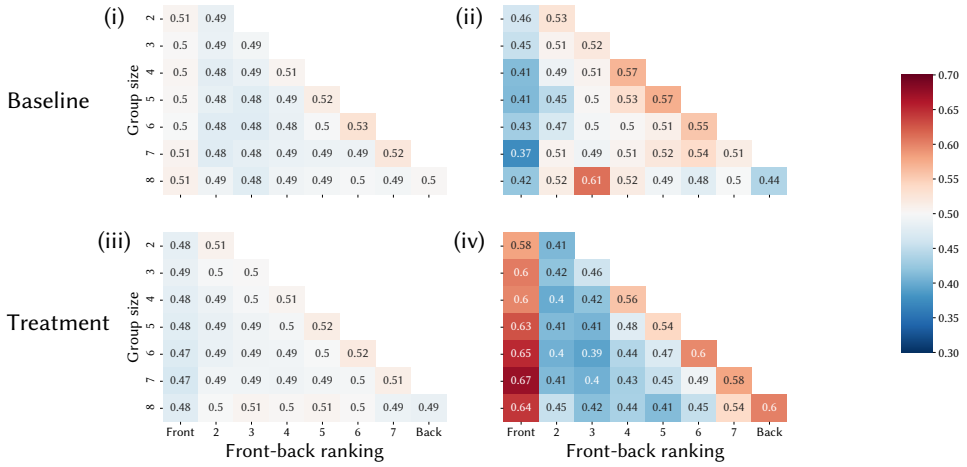
2.3.4 Speed and group size distributions

Relative speed, parasite load and infection duration

We compared swimming speed relative to the group mean for both the Focal-fish (FF) and the Non-focal fish (NF) in Baseline and Treatment conditions. The relative speed was calculated for each fish with respect to the trial mean speed, at each group size; the

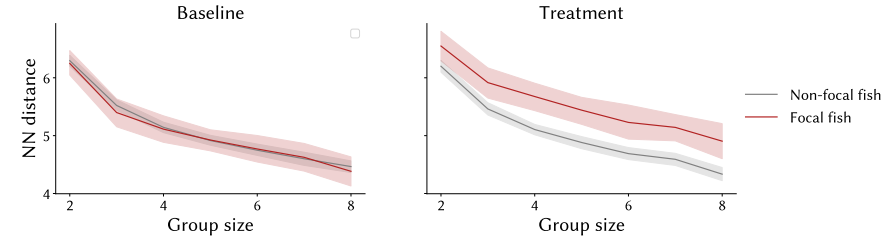


(a) Center-periphery distribution

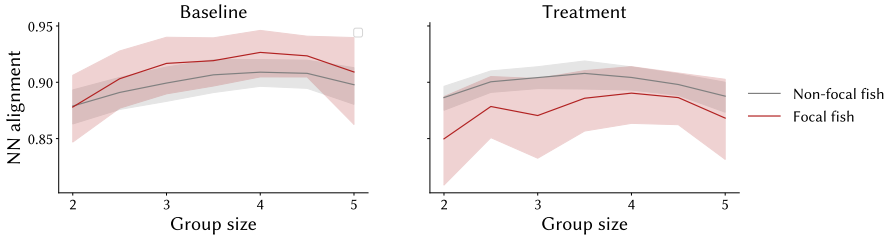


(b) Front-back distribution

Figure 2.10: **Ranked spatial position in the group**. (a) Normalized proportion of time in center-periphery positions in the group. The x-axis shows ranking positions, where rank = 1 is center and rank > 1 is periphery. Top row: Baseline; bottom row: Treatment. Comparison of the values of the proportion of time spent in each ranking, between the Non-focal fish (left column) and the Focal-fish (right column). Focal-fish in the Treatment condition spent more time in the periphery of the group (red values in ranking 2,3,4 for the group sizes of 4,5,6,7,8). (b) Proportion of time in front-back positions in the group. The x-axis shows ranking positions, where rank = 1 is front and rank > 1 is towards the center and back, depending on the group size. Top row: Baseline; bottom row: Treatment. Comparison of the values of the proportion of time spent in each ranking, between the Non-focal fish (left) to the Focal-fish (right). Focal-fish in the Treatment condition spent more time in the front of the group (red values in ranking 1 for all group sizes) and in the back (rad values in ranking 6,7,8 for the corresponding group sizes). All values are normalized by the group size and therefore a probability around 0.5 denotes little to no preference for a ranking position at any group size. The normalization is done as $(P_{ranking} * groupsize)/2$, therefore if the proportion time spent in ranking 1, of a group size of 8 is equal to 1/8 (random), then the normalized value will be equal to (0.125 * 8) / 2 = 0.5.



(a) Nearest-neighbor distance

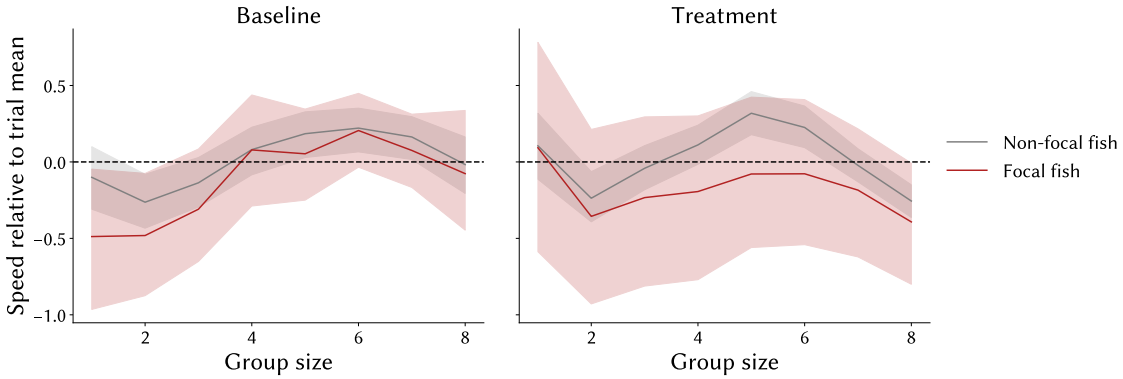


(b) Alignment to the nearest-neighbor

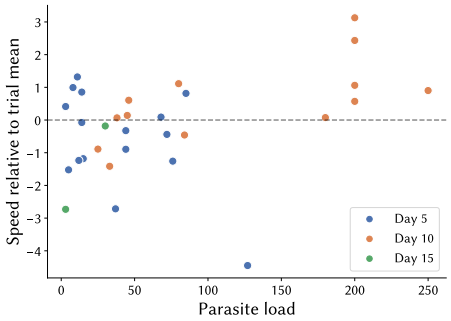
Figure 2.11: **Individual differences on nearest-neighbor distance and nearest-neighbor alignment with respect to group size.** (a,b) Left column shows data from the Baseline condition, right column shows data for the Treatment condition. Gray lines show pooled data of Non-focal fish, red lines show Focal-fish data. (a) In the Baseline condition (left) we find little to no difference between the Non-focal and the Focal fish. In the Treatment condition (right) we find that parasitized Focal-fish individuals (red) measure higher nearest-neighbor distance, across group sizes than the non-parasitized Non-focal fish (gray). See also Figure B.1 for the same result on the sham-infected fish in the control experiment. (b) We find little to no change in the nearest-neighbor alignment between Non-focal and focal-fish, in neither the Baseline (left) nor the Treatment (right) condition. Line plots show mean and 95% CI around the mean.

trial mean at each group size was calculated as the average of all fish in the trial. We find little to no differences in the relative speed between the NF and the FF in either of the experimental conditions, but we see a large variation in the relative speed as shown by the large CI in the Focal-fish Treatment condition (Figure 2.12a, Treatment). This large variation suggests that it's likely that the effect of parasitism on speed is not unidirectional, with some infected individuals moving either faster or slower than the mean. Therefore, we looked more in detail to understand what influences this variation. We find that relative speed is positively correlated with front-back distance, *i.e.* individuals that are faster than the shoal members in the same trial, are also more likely to be in the front of the group ($r=0.56$, $p < 0.001$), which suggests that spatial distribution is also determined by individual speed (Figure B.9). Figure B.9 shows the correlation between the front-back distance and the relative speed in the Baseline condition (left, $r = 0.5$, $p < 0.001$), compared to the Treatment condition (right, $r=0.62$, $p < 0.001$), for both Non-focal fish (gray) and Focal-fish (red).

In addition to the relationship to front-back distance, we looked at how the speed relative to the trial mean varies as a function of parasite load. We find that fish with high parasite load, on average swim at higher speed relative to the trial mean (Figure 2.12b).



(a) Speed relative to trial mean across groups sizes and treatments



(b) Relative speed as a function of parasite load and infection duration

Figure 2.12: **Speed relative to trial mean and relationship with other metrics.** (a) Speed relative to the trial mean across group sizes. We don't find differences between the Non-focal fish and the Focal-fish relative speed values neither at the Baseline (left) nor at the Treatment (right) condition. However, we see a large variation in the relative speed as shown by the large CI in the Focal-fish Treatment condition (right plot). Each line plot shows mean and 95% CI around the mean. (b) Correlation between parasite load and individual speed relative to the trial mean; values for the Focal-fish in the Treatment condition. We find that fish with high parasite load, on average swim at higher speed relative to the trial mean. The dotted line at zero denotes the trial mean. Dots are colored based on the day of the experiment (blue: 5, orange: 10, green: 15).

Time in different group sizes, role of speed

In this section, for each fish in each condition tested, we quantified the proportion of time each fish spent in each of the different group sizes. The proportion of time spent in each group size is normalized and calculated as $(f-m)/m$ where f is an individual and m is the mean of all the fish in the trial. We compare Baseline to Treatment, distinguishing focal from Non-focal fish. Data from Non-focal fish are pooled together and the confidence interval of the mean of the proportion of time spent in each group size is calculated across

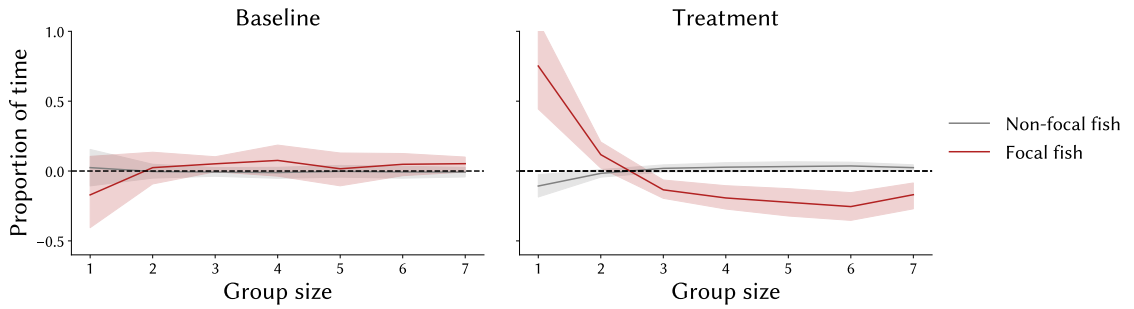


Figure 2.13: **Normalized proportion of time spent in different group sizes.** Comparison between the normalized proportion of time Non-focal fish (gray) and Focal-fish (red) spend in the different group sizes at the Baseline (left) and at the Treatment (right) condition. Data from Non-focal fish is pooled together and the confidence interval of the mean is calculated across fish. In the Baseline condition, we see little to no change in the proportion of time NF and FF spend in different group sizes. In the Treatment condition, the Focal-fish shows an increase in the time spent in group sizes of 1 (75% more than the trial mean), compared to the NF. Each group size is normalized and calculated as $(f - m)/m$ where f is an individual and m is the mean of all the fish in the trial. See also Figure B.2 for the same result on the sham-infected fish in the control experiment. Line plots show mean and 95% CI around the mean.

fish. Figure 2.13 shows how the distribution of time in the group sizes varies between Non-focal and Focal-fish at the Baseline (left) and at the Treatment (right). Focal-fish in the Treatment condition spend more time in isolation (75% more than the trial mean) compared to the Non-focal fish (17% less than the trial mean) (Figure 2.13). In order to verify that changes in the normalized proportion of time spent in isolation were not driven by effects of the manipulation on the fish in the experimental procedure, we compared these results to a control experiment (Appendix B.1), where the Focal-fish was sham-infected (Appendix B.1.1). We found that the proportion of time spent in isolation did not increase as a consequence of the sham-infection manipulation on the Focal-fish in the treatment condition (Figure B.2).

Besides the normalized proportion of time spent in each group size, we also compared the average duration (in seconds) of interactions at each group size, comparing Non-focal and Focal-fish across Treatments. Figure B.8 shows event duration is similar in the Baseline and in the Treatment. The only difference is found in the duration of the events in group size one in the Treatment condition: in this case, Focal-fish tend to have longer-lasting periods of isolation, than do Non-focal fish. On average the duration of isolation events for Focal-fish in the Treatment condition is 10.6 seconds, compared to 8 seconds for the Non-Focal fish (Figure B.8).

In order to determine a possible mechanism that explains isolation, we looked at the relationship of normalized proportion of time in isolation, with other behavioral metrics. We first looked at the relationship between the swim speed of each fish when they were

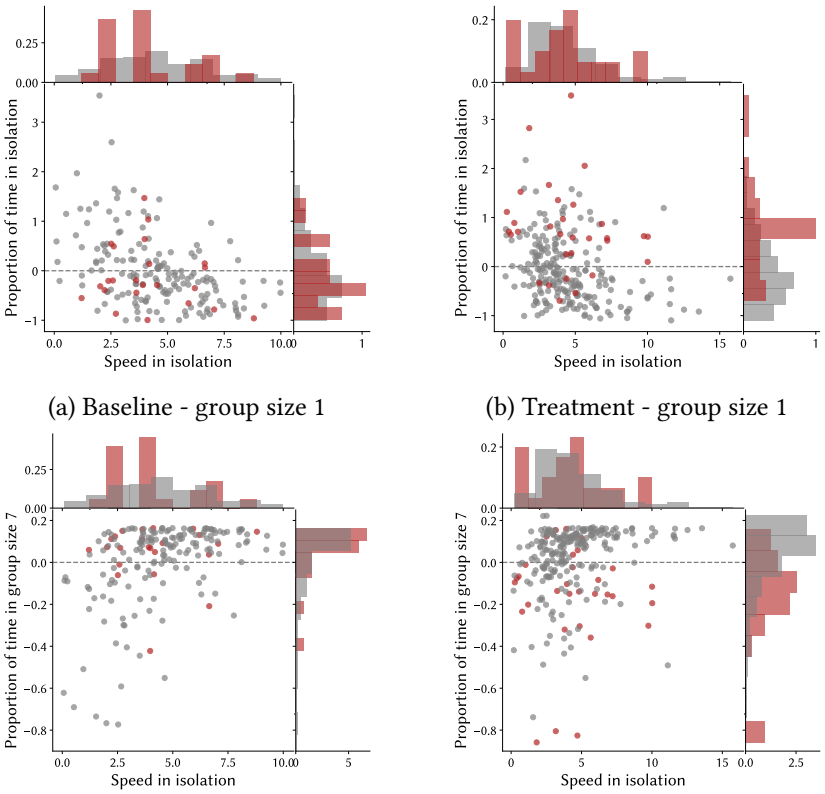
isolated (in the group size of one) - a measure of their 'preferred' swimming speed - and the time spent in isolation, as well as the time spent in a group size of 7 relative to the trial mean. We use the group size of 7 because it reflects a case where only one individual is isolated (Figure 2.14). Overall, the swimming speed when isolated is negatively correlated with the proportion of time spent in isolation ($r=-0.33$, $p < 0.001$), suggesting that slow individuals on average spend more time in isolation. We further compared this relationship between the Baseline (Figure 2.14a) and the Treatment (Figure 2.14b) condition. Both conditions show a negative correlation ($r=-0.44$, for the Baseline, $r=-0.27$ for the Treatment, $p < 0.001$), with an increased likelihood for the Treatment Focal-fish to be isolated at lower absolute swim speed in isolation (Figure 2.14b). Values of speed in isolation and the correlation with the time spent in isolation are similar between the Focal-fish and the Non-focal fish. Therefore, we conclude that the increase in the time spent in isolation of the Focal-fish cannot be explained by preferred speed only.

Next, we looked at how speed when isolated is correlated with the proportion of time each individual spends in the group size of 7. Overall, across conditions and fish, faster fish spend more time in the group size of 7 ($r=0.29$, $p < 0.001$). More specifically, the correlation between the two metrics is stronger in the Baseline condition ($r=0.43$, $p < 0.001$, Figure 2.14c) than it is in the Treatment condition (Pearson Correlation Coefficient $r=0.2$, $p < 0.001$, Figure 2.14d). Overall, we conclude that slower fish are both more likely to be isolated and less likely to be in a group size of seven.

2.3.5 Dynamics of group size changes

Guppies are commonly referred to as living within a fission-fusion population, which reflects group movements characterized by dynamic reshaping of the network (Wilson et al., 2014). As previously seen, in our data, with our definition of events, we find that group composition changes on average, and across group sizes, every 6.2 ± 2.8 s (Figure 2.8a, right). Until now, we have looked only at average duration and average proportion of time spent in different group sizes. Here, to further quantify aspects of the dynamical changes of shoaling guppies, we looked at the transition probability from a certain group size to another. In this context, we use the term 'transition' to refer to any fission or fusion event, where a fish leaves a group size to join the next group. Figure 2.15a shows an example where a group of 5 individuals divides into two smaller groups of 1 and 4 individuals (fission). To generalize the analysis, we identified all the transitions each individual undergoes throughout a trial and calculated the probability associated with the transition.

Starting with the Focal-fish, we first looked at the probability for the fish to transition from any group size (2 to 8) to the group size of one, that is the transition to isolation.



(a) Baseline - group size 1

(b) Treatment - group size 1

(c) Baseline - group size 7

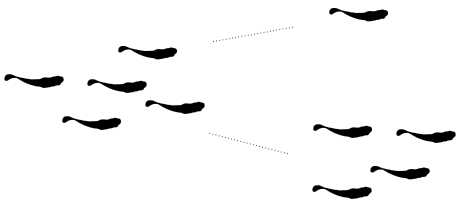
(d) Treatment - group size 7

Figure 2.14: **Correlation speed and isolation.** (a-b) Correlation between speed when isolated (cm/s) (group size of 1, x-axis) and normalized proportion of time spent in a group size of 1 (y-axis). Both conditions (Baseline left, Treatment right) show a negative correlation (r=-0.44 for the Baseline, r=-0.27 for the Treatment, $p < 0.001$), with an increased likelihood for the Treatment Focal-fish to be isolated at lower absolute swim speed (c-d) Correlation between speed when isolated (group size of 1, x-axis) and proportion of time spent in a group size of 7 (y-axis). Slower fish spend less time in the group size of 7, and the correlation is stronger in the Baseline condition (r=0.43, $p < 0.001$) compared to the Treatment condition (r=0.2, $p < 0.001$). The correlation coefficient was calculated for all fish in each condition without distinction between FF and NF. Scatter plots show one value per fish and histograms on each correspond to Non-focal fish (gray) and Focal-fish (red). The horizontal dotted line at zero represents the trial mean, therefore negative values indicate less time than average in that group size (and the opposite for positive values). The proportion of time on the y-axis is normalized and calculated for each individual. All correlation values are calculated with with Pearson Correlation Coefficient.

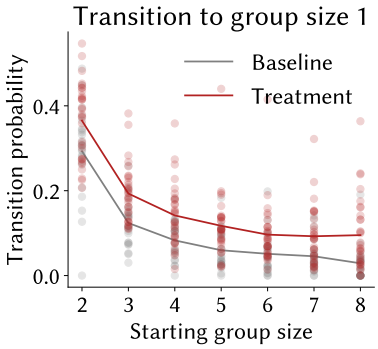
Figure 2.15b shows the transition probability from any group size (x-axis) to the group size of 1, comparing Baseline (gray) to Treatment (red). On average, in the Baseline condition, the probability for the Focal-fish to transition to isolation is 10%, compared to 16% in the Treatment condition (Figure 2.15b, means of the gray and red line across group sizes). In comparison, we also looked at the transition probability from the group size of 8 to any other group size (except 8) (Figure B.10). This comparison also shows how, on average, the Focal-fish in the Treatment condition are more likely to transition from group size 8 to a group size of 1 and 2 than Focal-fish are in the Baseline condition (Figure B.10). Finally, Figure 2.15c shows data of the fission-fusion associated probability of transitions between any starting group size, to any other group size (the next group to join). The left matrix of Figure 2.15c shows data for the Baseline condition while the right matrix of Figure 2.15c shows data for the Treatment. The first column of each matrix in Figure 2.15c describes the probability of transitioning from any group size (2-8) to the group size of one, previously also explained with reference to Figure 2.15b. In particular, we want to highlight the difference in the probability of being part of a fission from a group size of 8 to a group size of 1, which is equal to 3% of the time the Focal-fish in the Baseline condition, and to 10% for the Focal-fish in the Treatment condition (Figure 2.15c, bottom left corner). Finally, we compared these findings to the Non-focal fish (Figure B.11). Overall, we find little to no difference in the probability for the Non-focal fish to transition from any group size to a group size of one when comparing the Baseline and the Treatment condition (Figure B.11b), as well as to all other group sizes (Figure B.11c).

As a further note, we find that in the guppy movement dynamics, a change in group composition is mainly driven by one fish either leaving or joining a group, as highlighted by the dark diagonal in Figure 2.15c (true also for Non-Focal fish, see Figure B.11). In other words, when we define a fission or a fusion in our guppy system, it is likely that the group size change is driven by the move of one individual, that either joined or left the group. An obvious example is the previously mentioned case, where one individual is separated by the group of 8, and thus the group size of 8 changes to the group size of 7. In this case, the transition 8-to-7 is the most likely of all transitions, which for the Focal-fish happens in 37% and 30% of the cases in the Baseline and Treatment conditions, respectively.

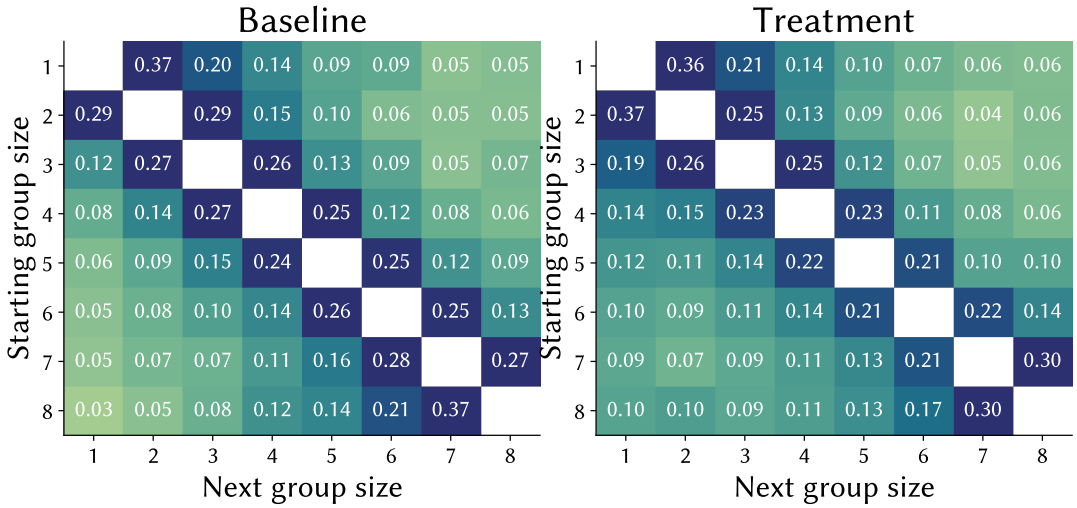
In addition to the analysis above, where we looked at the probability of each fission and fusion event for each group size, here we describe the temporal aspect of fission/fusion event. More specifically, for each individual we looked at how the group size varies in the 33 seconds that follow a fission or a fusion event. The motivation for this analysis comes from the observation that groups within our population vary fre-



(a) Example transition from group size of 5 to groups of 4 and 1



(b) Transition probability 8 to 1



(c) Transition matrix for all combinations

Figure 2.15: **Fission-fusion transition probability for the Focal-fish.** (a) Example of fission from a group size of 5 to two smaller groups of 1 and 4 individuals. (b) Transition probability of the Focal-fish from any group size (x-axis) to the group size of 1, in the Baseline (gray) and the Treatment (red) condition. Lines show the mean transition probability for the Focal-fish from any group size (2-8, x-axis) to the group size of 1. The dots show data for each Focal-fish in the Baseline (gray) and in the Treatment (red) condition. (c) Fission-fusion associated probability for the Focal-fish of transitioning between any starting group size (y-axis), to any other group size (the next group to join, x-axis). The first column of each matrix refers to the probability of the Focal-fish at the Baseline (left) and at the Treatment (right) to transition from any group size (2-8) to the group size of 1, as shown in 2.15b. Color denotes transition probability, from low (light) to high probability (dark).

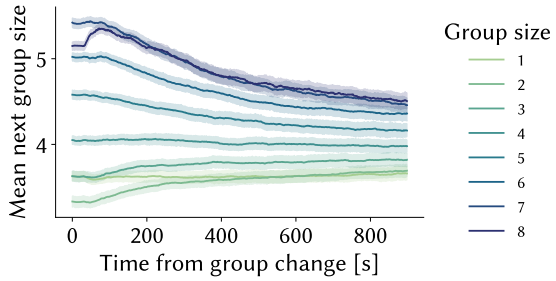
quently, and fission and fusion events can have cascading effects in rapidly minimizing or maximizing, over time, the group size a fish is in. On average across all trials, guppies are more likely to change group size by one unit (*e.g.* from a group size of 7 to either a group size of 6 or 8) (Figure 2.16a, also visible in Figure 2.15c, dark diagonal). Next,

to simplify the quantification, we focus on a clear fusion event (transition from a group size of 1 to any other group size (Figure 2.16b) and on a clear fission event (from a group size of 8 to any other, Figure 2.16c), and compare the Baseline to the Treatment for both the Non-focal fish (gray) and the Focal-fish (red). Figure 2.16b shows the average across Baseline (left) and Treatment (right), for the Non-focal fish (gray) and the Focal-fish (red), of the group size over the 33 seconds following a fusion event from the group size of 1 to any other group size. In this comparison, in the Baseline condition, the confidence intervals around the mean are overlapping and we therefore do not find any difference between the Non-focal fish (gray line) and the Focal-fish (red line) (Figure 2.16b, left); we note that the Focal-fish shows a very large confidence interval around the mean. On the other hand, in the Treatment condition, we find that the Focal-fish, in the 33 seconds following a fusion is found progressively in smaller groups (Figure 2.16b, right, red line). Finally, Figure 2.16c shows the average group size over the 33 seconds following a fission event from the group size of 8 to any other group size. Similar to what we found for the data after fusion events, in the Baseline condition, the confidence intervals around the mean are overlapping and we therefore do not find any difference between the Non-focal fish (gray line) and the Focal-fish (red line), which means that following a fission, the mean group size over time does not significantly differ between the NF and the FF (Figure 2.16c, left). In the Treatment condition, we find that following a fission event from a group size of 8, on average, the Focal-fish is more often found in smaller group sizes than the Non-focal fish, confirming our previous results (Figure 2.16c, right). For both the NF and the FF, following fission from a group size of 8, the average group size progressively decreases over time to reach the respective mean value (Figure 2.16c, right).

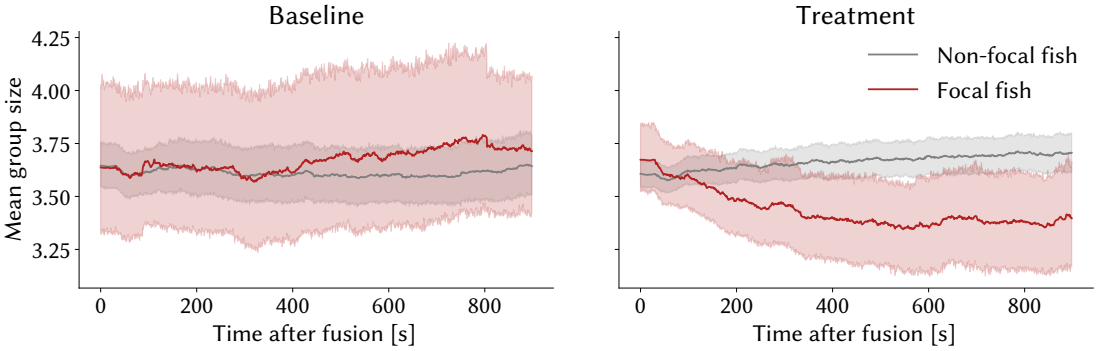
Metrics correlated to the increased likelihood of fission to isolation

In the previous section, we saw that a fish, when parasitized, is more likely to transition to being isolated from a group size of 8. In this section, we look at what behavioral predictors might explain the increased likelihood of the transition. The metric of transition to isolation is different from the proportion of time spent in isolation because it captures a dynamic change. In the extreme case, a fish could spend all of its time in isolation without ever transitioning from any other group size.

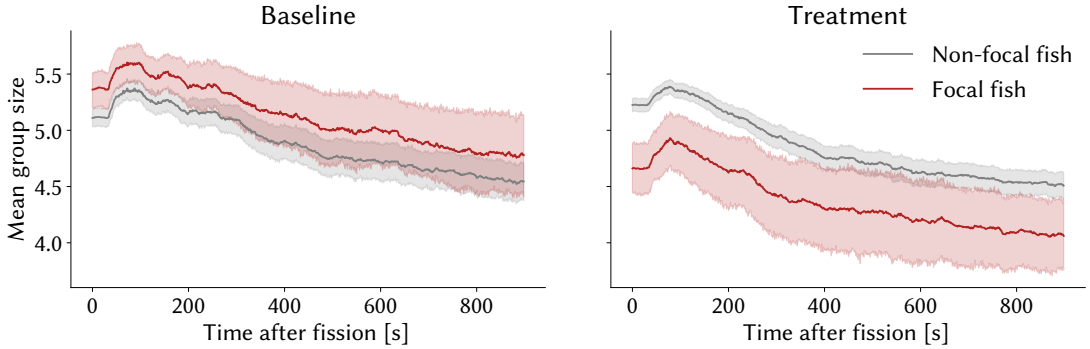
The first correlation we look at for each fish, is speed, in this case as in mean speed relative to the trial mean. Figure 2.17a shows the correlation between the speed of all fish in a group size of 8 relative to the trial mean, and the transition probability from a group size of 8 to a group size of 1 (8-to-1 fission). The left side of Figure 2.17a shows the data for the Baseline condition, where we see a general negative but small correlation between



(a) Average group size after fission



(b) Average group size over time after fusion from group size 1



(c) Average group size over time after fission from group size 8

Figure 2.16: **Average group size over the 33 seconds following a fission or a fusion event.** (a) Average group size over time (x-axis) after a fission or a fusion event; each line shows the group size a fish is in before a fission or a fusion. On average, a fish transitions to a group that is similar in size (y-axis). (b) Average group size over the 33 seconds following a fusion from a group size of 1 to any other group size. (c) Average group size over the 33 seconds following a fission from a group size of 8 to any other group size. Comparison between the Non-focal fish (gray) and the Focal-fish (red) in the Baseline (left) and the Treatment (right) condition, respectively. Line plots show mean and 95% CI around the mean.

the relative speed of a fish, and the 8-to-1 fission probability, for both the Non-focal and the Focal-fish (r=-0.25, p=0.001 and r=-0.16, p>0.05 respectively). The right side of Figure 2.17a shows that in the Treatment condition, the Focal-fish shows a relatively stronger negative correlation between speed and probability of transition than the Non-focal fish in the same condition (r=-0.45, p=0.01 and r=-0.24, p<0.001, respectively). These results confirm what was previously shown in section 2.3.4, where we showed that the fish with

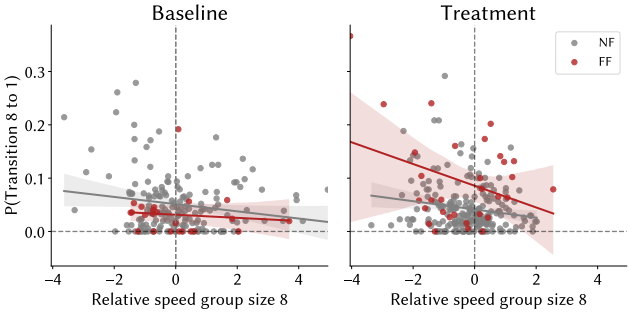
lower preferred speed (as measured when isolated) are more likely to be isolated. Here, we show that the transition probability from a group size of 8 to isolation is also possibly driven by differences in speed relative to the trial mean.

Following the correlation with speed, we looked at how the probability of the transition to isolation correlates with a fish's turning rate. In this case, we look at the average turning rate a fish shows in isolation, since when in groups the turning rate is a function of both a focal individual's intrinsic tendency, but also the complex social environment in which it is moving. Figure 2.17b shows how turning rate when isolated correlates with the 8-to-1 fission probability, in the Baseline (left) and in the Treatment (right) condition, comparing Non-Focal fish (gray) to Focal-Fish (red). For the Baseline condition, we do not find a correlation between turning rate and 8-to-1 fission probability, for both the Non-focal fish and the Focal-fish (Figure 2.17b, left). In the Treatment condition, the Non-focal fish also shows no correlation between the turning rate and the 8-to-1 fission probability (Figure 2.17b, right, gray line), whereas the Focal-fish shows a relatively strong negative correlation (r=-0.42, p = 0.02) (Figure 2.17b, tight, red line). This suggests that it's possible that changes in turning rates when isolated might partly explain the increase in the likelihood of transition to isolation for the Focal-fish in the Treatment condition.

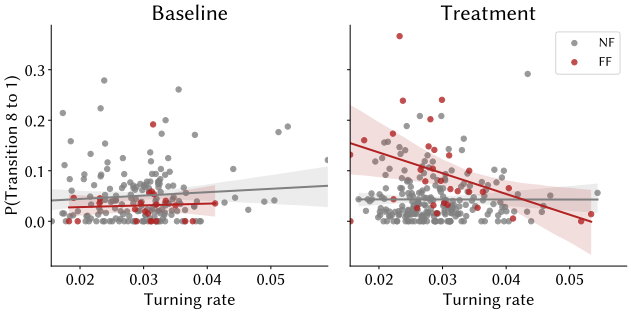
Finally, we find that parasite load is positively correlated with increased probability in the transition from group size 8 to isolation (r=0.47, p<0.001, Figure 2.17c, left). Differently, parasite load is weakly negatively correlated with turning rate in isolation, although the correlation is not significant (r=-0.23, p>0.05, Figure 2.17c, right). This indicates that increasing infection loads increases the likelihood of a fish being part of fission events that lead to an isolated state.

2.3.6 Body size, parasite load and behavioral metrics in relation to parasite transmission

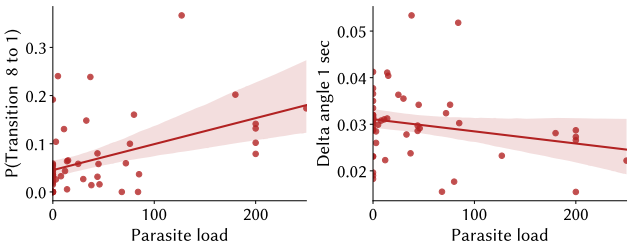
The following section explores potential predictors of parasite transmission, both from the perspective of the infected fish, that transmits infection and from the perspective of the (yet) uninfected fish, that can get infected during a trial when swimming in proximity of the infected Focal-fish. For the following analysis, we exclude fish that were tested more than one time in the Treatment condition to avoid possible familiarity effects with the tank and the experimental setup, that could have affected the transmission rate (for example, facilitating transmission on the second time a fish is tested). We therefore include only 25 Treatment trials (compared to the 31 used for the previous analysis). We recorded parasite transmission in 9 out of 25 trials, of which 18.7% happened on the day



(a) Transition to isolation and relative speed



(b) Transition to isolation and turning rate



(c) Transition to isolation, turning rate and parasite load

Figure 2.17: **Behavioral predictors of the probability of fission to isolation.** (a) Correlation between the speed of a fish in a group size of 8 relative to the trial mean and the transition probability from a group size of 8 to a group size of 1. Each plot compares the NF (gray) and the FF (red) between the Baseline (left) and the Treatment (right) condition. In the Baseline condition, we see a general negative but small correlation for both the Non-focal fish and the Focal-fish (r=-0.25, p=0.001 and r=-0.16, p>0.05 respectively). In the Treatment condition, the Non-focal fish shows a similar trend to the Baseline, whereas the Focal-fish shows a relatively stronger negative correlation (r=-0.24, p<0.001 r=-0.45, p=0.01, respectively). (b) Correlation between turning rate and transition probability from a group size of 8 to a group size of 1. In the Baseline condition, we see little to no correlation for both the Non-focal fish and the Focal-fish. In the Treatment condition, the Non-focal fish shows again little to no correlation, whereas the Focal-fish shows a relatively strong negative correlation (r=-0.42, p = 0.02). (c, left) Correlation between parasite load and transition probability from a group size of 8 to a group size of 1. Parasite load is positively correlated with increased probability in the transition to isolation (r=0.47, p<0.001). (c, right) Correlation between parasite load and turning rate in isolation. We find decreasing turning rates in isolation at increasing parasite load, although the correlation is not significant (r=-0.23, p>0.05). Data in this plot include all Focal-fish in both Baseline and Treatment condition. Line plots show mean and 95% CI around the mean.

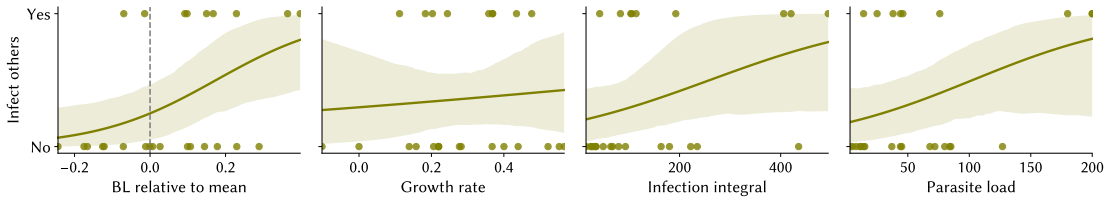
5 of testing, and 66% happened on the day 10 of testing. In trials where parasites were transmitted, the Focal-fish infected a number of Non-focal fish between 1 and 4. More specifically, in 6 trials we counted 1 fish infected, in 2 trials we counted 2 fish infected, in 1 trial we counted 4 fish infected.

First, from the infected Focal-fish's perspective, in order to explore what factors may predict infection transmission, we fitted a logistic regression model to both 'physiological' and behavioral metrics, separately. As physiological metrics, we included the body size of the Focal-fish relative to the trial mean, parasite growth rate, parasite load infection integral, and absolute parasite load (number of worms at day of testing). Behavioral metrics included speed relative to the trial mean, distance from the centroid, nearest-neighbor distance, and proportion of time spent in isolation. First, we fitted a logistic regression model to the physiological metrics.

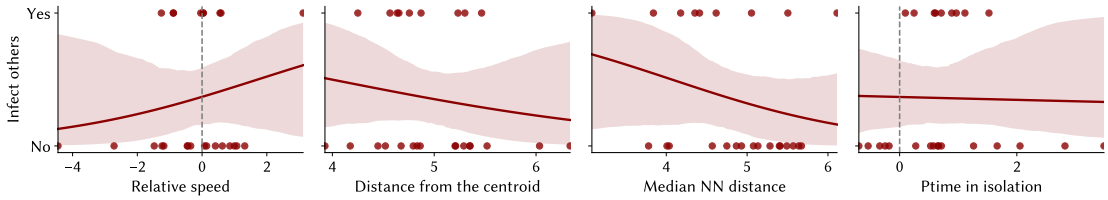
Logistic regression was used to analyze the relationship between the body size of the Focal-fish relative to the trial mean, parasite growth rate, parasite load infection integral, absolute parasite load, and infection transmission likelihood. We found that holding all other predictor variables constant, the odds of infection transmission likelihood occurring increased by 21% (95% CI [0.032, 0.389]) for a one-unit increase in body size of the Focal-fish relative to the trial mean (p=0.021). Moreover, the odds of infection transmission likelihood occurring increased by 41.4% (95% CI [0.029, 0.799]) for a one-unit increase in the Focal-fish parasite load (p=0.029). Figure 2.18 shows the line fit to the input variables of the logistic regression model. To evaluate the results we used Average Marginal Effects (AMEs) (Leeper, 2017). These results suggest that individuals that are bigger than the group average and that have higher parasite loads are more likely to have higher transmission rates.

Second, we used logistic regression to analyze the relationship between speed relative to the trial mean, distance from the centroid, nearest-neighbor distance, proportion of time spent in isolation, and infection transmission likelihood. Figure 2.18 shows the line fit to the behavioral input variables of the logistic regression model. The model shows no significant results. We conclude that these metrics are insufficient to predict infection transmission likelihood and that it is possible that more information at the level of the individual is needed.

Next, we looked at the transmission likelihood from the perspective of the Non-focal fish that became infected during a trial, and we explored which predictors might increase the likelihood of getting infected. We used logistic regression to analyze the relationship between the parasite load of the Focal-fish, the size of the Non-focal fish relative to the Focal-fish and the percentage of time spent close to the Focal-fish (closer than 3 body lengths) to predict the likelihood of getting an infection during the trial.



(a) Physiological predictors of transmission



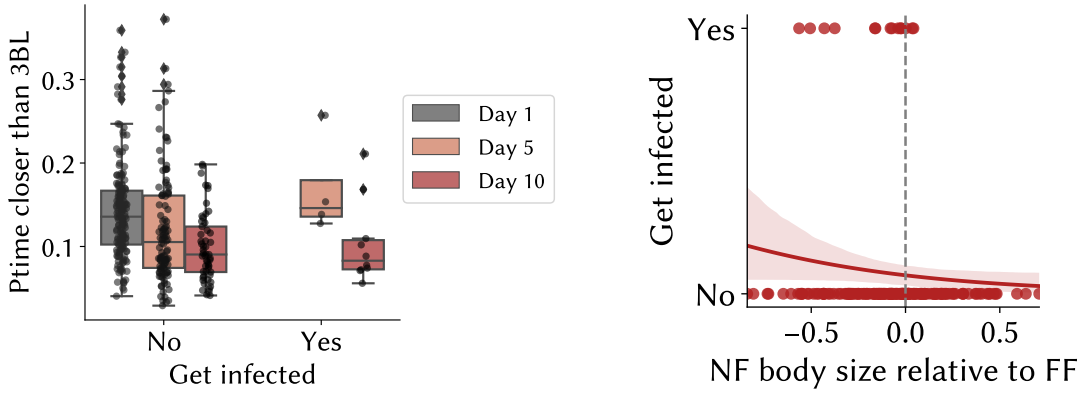
(b) Behavioral predictors of transmission

Figure 2.18: **Physiological and behavioral predictors of parasite transmission in the trial.** Logistic regression fit for four 'physiological' (a) and behavioral (b) metrics used as predictors of parasite transmission (y-axis). (a) Logistic regression fit for four physiological metrics used as predictors of parasite transmission (y-axis). We used body length of the Focal-fish relative to the trial mean, growth rate, infection integral, and parasite load as predictors. (b) Logistic regression fit for four behavioral metrics used as predictors of parasite transmission (y-axis). As behavioral predictors we used speed relative to the trial mean, distance from the centroid, nearest-neighbor distance (cm), and proportion time spent in isolation. None of these is a significant predictor for infection transmission. Line plots show the mean and 95% CI of the mean.

We found that, holding all other predictor variables constant, the odds of the likelihood of getting infected during a trial increased by 6% (95% CI [0.022, 0.098] for a one-unit increase in parasite load of the Focal-fish (p=0.001). Similarly, we found that, the odds of the likelihood of getting infected increased by 4% (95% CI [0.003, 0.082] for a one-unit increase in the percentage of time spent in close proximity to the Focal-fish (p=0.036).

Figure 2.19 shows two of the predictors of the model, that is time spent in close proximity to the FF and NF relative body size. In Figure 2.19a we show the proportion of time spent closer than 3 body lengths to the Focal-fish as a function of the day of the trial, which indicates the duration of the infection on the Focal-fish. We find that the time spent close proximity to the FF seems to be more relevant on short infection durations (day 5) than it is for the long infection durations (day 10) (Figure 2.19a). This suggests that the likelihood of getting infected for non-parasitized fish might depend on other factors than average proximity and that the increased load on day 10 of the experiment might influence infection transmission differently than on day 5 of infection. Figure 2.19b shows how the probability of the NF to getting infected during the trial correlates with respect to its size relative to the FF. Therefore, individuals that spend proportionally more time in close proximity to the infected Focal-fish, are more likely to get infected

during the trial. Moreover, higher parasite loads of the Focal-fish increase the likelihood of the Non-focal fish getting infected. We find no significance for the body size relative to the FF, which could be due to little data, given that only a small number of individuals got infected across all trials, compared to the total number of tested fish. Despite the lack of the significance of the model, we note that it was individuals smaller than the FF who tended to have contracted an infection during the experiment (Figure 2.19b),



(a) Infection likelihood and proximity to FF

(b) Infection likelihood and body size

Figure 2.19: **Predictors of receiving an infection.** Left: Proportion of time the Non-focal fish spent closer than 3 body lengths to the Focal-fish and likelihood, divided between fish that get and do not get infected during a trial. Proximity seems to play a role only for the short infection duration (day 5) and not for the long infection duration (day 10). Box plots give the median, interquartile range (box), and values that sit outside of 1.5 times the interquartile range (whiskers). Right: logistic regression fit on likelihood of getting infected as a function of size of the Non-Focal fish relative to the Focal-fish. The effect is not significant but there seems to be an indication that in our trials, only smaller fish had an increased likelihood of getting infected.

2.3.7 PCA on the Focal-fish behavioral metrics

Individual variation is a key component of animal behavior across taxa. In this section, we used Principal Component Analysis (PCA) to explore the variation of the behavioral metrics of the Focal-fish, to define gradients of behavioral phenotypes that define inter-individual variability that is associated with the infection status of the Focal-fish.

In the previous sections, we looked at individual metrics separately, comparing Focal- and Non-focal fish, between the Baseline and the Treatment conditions. We found many cases in which the Focal-fish behavior varies from the Baseline to the Treatment in several metrics, such as isolation rates, spatial ranking position in the group, relative speed, and nearest neighbor distance. Here, we used PCA to generalize the variation in the behavioral metrics of the Focal-fish as a function of infection status - which are calculated specifically for each group size. For this analysis, we included measures of relative speed,

nearest-neighbor distance, nearest-neighbor alignment, the proportion of time spent in the front of the group, the proportion of time spent in the periphery of the group, the probability to transition to isolation, and proportion of time spent in the different group sizes.

The first four components from the PCA decomposition of individual behavioral metrics are shown in Figure 2.20. All the weightings of each metric measured across group sizes, for the first four principal components, are shown in Figure 2.20. Positive/negative weightings in the PCA components represent higher/lower values of a metric with respect to the average. The first principle component explains the largest fraction of variance (35.8%), mainly weighted by the time spent in isolation, the peripheral position in the group, and the likelihood of transitioning to isolation. Positive projections of movement data onto the first PCA component therefore describe fish that spend more time in isolation, are more likely to be part of a fission to isolation, and are more peripheral in the group (Figure 2.20, first column). The second PCA component explains 16.3% of the total variance and is most strongly weighted by relative speed, nearest-neighbor distance, and proportion of time spent in the front of the group. Positive projections onto the second PCA component represent fish with higher relative speed, higher nearest neighbor distance, and higher proportion of time spent in the front of the group (Figure 2.20, second column). The third PCA component explains 10.7% of the total variance and is mainly weighted by nearest-neighbor alignment. Positive projections onto the second PCA component represent fish with lower nearest-neighbor alignment (Figure 2.20, third column). Finally, the fourth PCA component (9% of the total variance) is mainly weighted by nearest-neighbor distance and the proportion of time spent in the front of the group.

Figure 2.21a shows the projections onto the first two PCA components for the Focal-fish, comparing the Baseline to the Treatment. Values of the projections onto the first PCA component are mainly negative for the Focal-fish in the Baseline condition and positive for the Focal-fish in the Treatment condition (Figure 2.21a). Similarly, values of the projections onto the second PCA component are mainly negative for the Focal-fish in the Baseline condition, whereas values for the Focal-fish in the Treatment condition are both positive and negative (Figure 2.21a). Moreover, it looks like the points with the highest positive projection onto PC2 tend to be from the Focal-fish in the Treatment condition. Similarly, we compared projections onto PC1 to projections onto PC3 (Figure 2.21b). In this case, looking at the projections onto the third PCA component, we do not find a clear separation between the Focal-fish at the Baseline and Treatment conditions (y-axis, Figure 2.21b). However, we note that the fish with the highest positive and the few lowest negative projections onto PC3 are from the Focal-fish in the Treatment con-

dition (Figure 2.21b). Finally, variation among treatment fish on PC3 is higher for the Treatment fish than it is for the Baseline (Figure 2.21b). The values of the projection onto PC2 confirm the variability we previously described in section 2.3.4, where we find variation among Focal-fish in the treatment condition, with some individuals with distinct higher relative speed values, compared to other individuals with lower relative speed. Moreover, positive projections onto PC2 also show higher values of the proportion of time spent in the front of the group, which confirms that the same variability in relative speed will also mirror differences in the position in the group of the Focal-fish. Overall the first two principal components highlight the main separation between movement metrics of the Focal-fish in the Baseline condition, compared to those in the Treatment condition. Both PC1 and PC2 show a positive correlation with values of parasite load ($r=0.47$, $p<0.001$, both; Figure B.12). For a figure showing the 95% confidence interval around each Principal component, see Figure B.13.

2.4 Discussion

In this work, we used experimentally-infected guppies and identity-based tracking to examine how parasites affect shoaling behavior, and how the presence of a parasitized fish in a group impacts transmission and the motion of shoal members. Overall, we find that guppies infected with the ectoparasite *Gyrodactylus* sp. show behavioral differences at the individual level with respect to uninfected guppies, in the way they move and interact.

Guppies infected with *Gyrodactylus* sp. swim in more peripheral positions compared to uninfected fish (Figure 2.10a). Because they are more in the periphery, we also find that the infected fish are less likely to be the nearest-neighbors of uninfected fish (Figure 2.9b). We also see an increased tendency for infected guppies to occupy more frontal positions (Figure 2.10b), which is related to the swim speeds exhibited by fish swimming when isolated (*i.e.* in the same trial, but swimming apart from others), with faster individuals being more likely to occupy frontal positions (Figure B.9). In addition, infected individuals exhibited a greater proportion of their time isolated from others. This is related both to an increased duration of time spent isolated (Figure 2.13) and to an increased probability of being part of fissions from a large group size (8) to isolation (Figure 2.15). Both metrics, show an effect specific to the infected Focal-fish, and both rates increase at increasing parasite loads. Finally, we looked at what features of the infected Focal-fish correlated with the increased likelihood of transmitting the infection during a trial. On the other hand, we analyzed the features that increased the likelihood for an uninfected fish of becoming infected during the trial. From the perspective of the Focal-

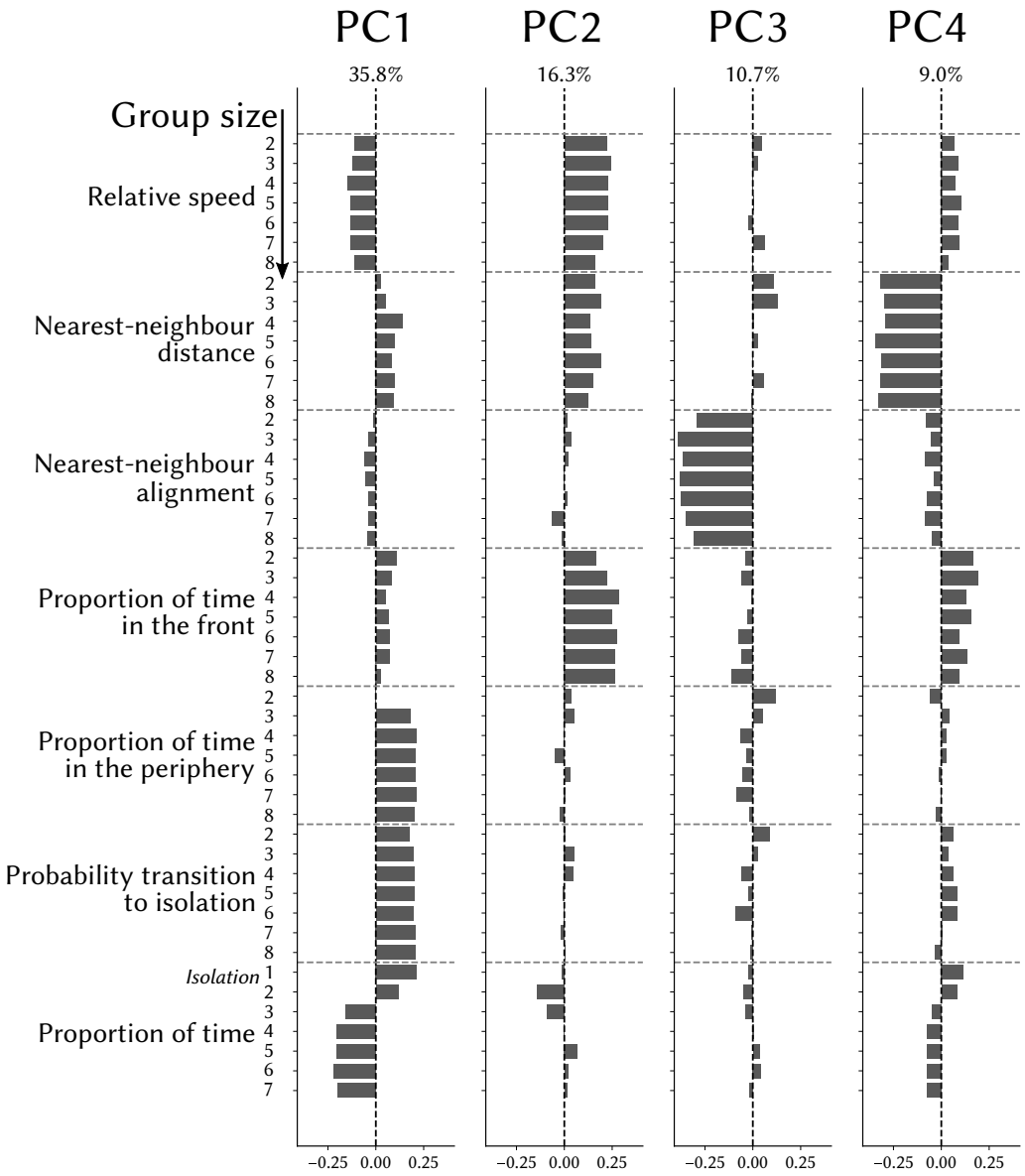
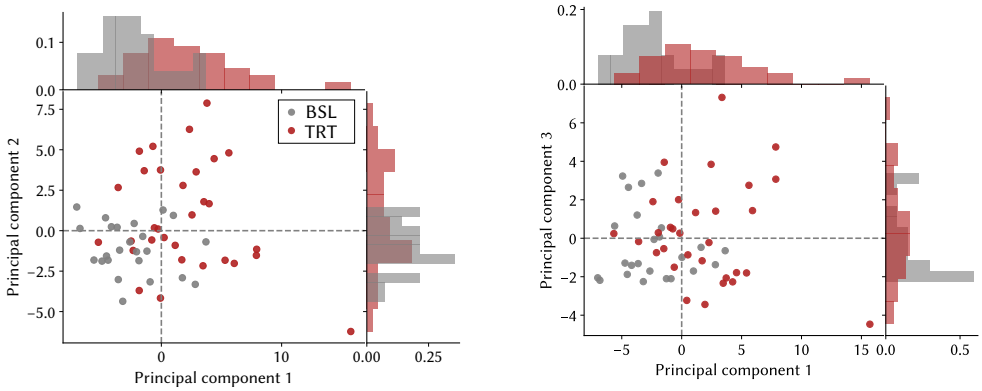


Figure 2.20: **The first four components of the PCA.** The first four components from the PCA decomposition of individual fish behavioral metrics represent the dominant axes of behavioral variation. Each value represents metrics of each fish in a range of group sizes (y-axis). Top left: positive and negative projections onto PCA1 and PCA2 represent different valence of the behavioral metrics of the PCA input.

fish transmitting an infection, we find that measures of relative body size and parasite load are the best predictors for infection transmission during a trial, with bigger fish with higher parasite loads being more likely to transmit parasites (Figure 2.18a). On the other hand, from the perspective of the uninfected fish that receive an infection during the experiment, we find that fish that spend more time closer than 3BLs to the Focal-fish are more likely to get infected during a trial, in particular, when in contact with fish at short (5 days) infection durations (Figure 2.19a).



(a) PC1 vs. PC2 (b) PC1 vs. PC3

Figure 2.21: **Projections onto the first 3 PCA modes, values for the Focal-fish.** (a) Projection onto PCA1 (x-axis) and PCA2 (y-axis) and (b) projection onto PCA1 (x-axis) and PCA3 (y-axis) for the Focal-fish in the Baseline (gray) and the Treatment (red). (a) Values of the projections onto the first PCA component are mainly negative for the Focal-fish in the Baseline condition and positive for the Focal-fish in the Treatment condition. Similarly, values of the projections onto the second PCA component are mainly negative for the Focal-fish in the Baseline condition, whereas values for the Focal-fish in the Treatment condition are both positive and negative. (b) Values of the projections onto the third PCA component (y-axis) do not show a clear separation between the Focal-fish in the Baseline and Treatment conditions. We note that the fish with the highest positive and the few lowest negative projections onto PC3 are from the Focal-fish in the Treatment condition.

2.4.1 Changes in group composition

Behavioral differences and variations of the internal state of an individual can affect the structure of the group (Couzin and Krause, 2003; Hoare et al., 2004; Ward et al., 2002). For example, the hunger state of fish increases nearest-neighbor distance (Morgan, 1988; Robinson and Pitcher, 1989a,9), predation risk increases shoal cohesion (Krause, 1993; Magurran et al., 1987), or larger individuals often occupy more frontal positions (Hoare et al., 1998), as well as food-deprived roaches (*Rutilus rutilus*) (Krause et al., 1992), where they probably benefit of increased feeding rates (reduced competition). In our study, we altered the infection state of one fish of a shoal of guppies, by experimentally infecting it with *Gyrodactylus* sp. and quantified how individual spatial positioning was affected by the infection status. Here we discuss results on changes in nearest-neighbor distance as well as changes in the position infected and uninfected fish take in the group for each individual.

In the analysis of the distribution of nearest-neighbors, and of nearest-neighbor distances, we find that the infected Focal-fish is less likely to be the nearest-neighbor of any other fish in the group, while it also shows, on average, higher nearest-neighbor distance than the Non-focal fish. These results together are consistent with the findings of Richards et al. (2010) where parasite prevalence increased distance between shoaling

guppies. Similarly, in a study that looked at the effect of water flow on shoaling and parasitism of guppies, it was also shown that in absence of water flow, shoals were less cohesive (Reynolds et al., 2019). Overall, reductions of shoaling cohesion in the presence of disease, were previously reported on several species (Dugatkin et al., 1994; Hockley et al., 2014; Krause and Godin, 1996), and often interpreted in light of possible avoidance mechanisms; however, we believe that other mechanisms such as self-organized sorting may also play a role (Couzin and Krause, 2003), as we will discuss more in detail later in this section. Alternatively, increased nearest-neighbor distance, in the absence of other group-level changes, such as disruptions of fission-fusion dynamics, could also be a mechanism to minimize phenotypic heterogeneity - and oddity effects, while decreasing the infection risks. For example, a study by Croft et al. (2011) previously showed the effects of parasitism on fission-fusion dynamics, finding that guppies shoal would initiate more fission events when an infected conspecific was in the group. Theory predicts that shoal members should make decisions that minimize oddity and maximize the dilution effect, in particular to avoid predation (Barber et al., 1998).

In addition to the analysis of nearest-neighbors distance, we quantified spatial positioning for each member in the shoal, ranking each individual position across group sizes and computing the likelihood associated with each rank. We find that infected fish are more likely to occupy peripheral positions, while also showing increased presence in frontal ranking positions (Figure 2.10). In the parasite literature, it was previously observed that parasitized banded killifish (*Fundulus diaphanus*) and minnows (*Phoxinus phoxinus*) occupied more peripheral positions when part of a shoal (Barber and Huntingford, 1996; Ward et al., 2002). In both cases, however, the studied parasites had complex life cycles, and this is hypothesized to increase selection pressure to impact the host's behavior to go to the next host (Poulin, 2000). Finally, we also find that infected guppies are more likely to be in the front of the shoal (Figure 2.10b), which was also previously observed in relation to altered nutritional states and to parasitism (Krause et al., 1992; Ward et al., 2002).

In light of these results and in the context of previous findings, we hypothesize that *Gyrodactylus* sp. infection possibly affects the internal state of the fish, which consequently alters the spatial positioning and collective behavior. One possibility is that those individuals that take frontal shoal positions are in a relatively more beneficial position for increasing feeding rates, compared to the other shoal members (Hoare et al., 1998; Krause et al., 1992; Ward et al., 2002). Individual differences of the shoal members can lead to self-organization sorting mechanisms: similar phenotypes naturally sort and associate, resulting in differences in spatial positioning (Couzin and Krause, 2003). Therefore, also in absence of complex mechanisms, sorting mechanisms may

help understanding of how infected individuals modify their position relative to others (Couzin and Krause, 2003; Couzin et al., 2002). Unlike other studies where the impact on host's behavior is likely a result of selection pressure related to parasites with complex life cycles (Poulin, 2000), in our study we see evidence of similar behavioral effects that are resultant of an ectoparasitic infection with a direct life cycle, where the behavioral response to the infection in this social animal is probably host-driven (Hawley et al., 2021; Stockmaier et al., 2021). Finally, it is possible that the small ratio of infected-to-uninfected individuals in this study did not allow the emergence of major group-level changes.

2.4.2 Increased isolation rates

In the previous section, we discussed how an individual's position within the group changes following parasitic infection. Here, we will discuss more how spatial positioning is also seen in association with increases in isolation rates, and how differences in fission-fusion dynamics and isolation rates are related to the infection status of the individuals.

The isolation status of an individual is, in other words, distancing. The word distancing has received much attention lately, in particular in the meaning of 'social distancing', *i.e.* the reduction of social contacts in the presence of an infectious disease or pathogen, which is found in both animal and humans (*e.g.* Moreno et al. (2021); Stockmaier et al. (2021)). 'Social distancing' however is not the only behavioral response that animals (and possibly humans) have adopted to avoid or fight infectious diseases. Briefly, among the main 'direct' behavioral responses that social species show upon exposure to pathogens, we find passive and active self-isolation (host-driven), avoidance, exclusion, and group-wide social distancing (driven by uninfected conspecifics). Among self-isolation behaviors, we also find 'indirect' behavioral responses, where the infected host for example reduces social contacts due to a physiological response such as lethargy or loss of interest in social interactions (for a comprehensive review, see (Stockmaier et al., 2021). The important difference between active and passive self-isolation is that active self-isolation prevents interaction with conspecifics, whereas passive self-isolation does not (Stockmaier et al., 2021). These behavioral responses vary in the implications they have on infection transmission, by altering, for example, social connectivity (Lopes et al., 2016; Ripperger et al., 2020; Stroeymeyt et al., 2018).

Guppies have been previously shown to adopt the strategy of active avoidance of infected conspecifics (Stephenson, 2019; Stephenson et al., 2018). This strategy has been shown to vary with respect to the costs of the infection, with highly susceptible individuals showing stronger avoidance (Stephenson, 2019), or when transmission risk is higher (Stephenson et al., 2018). Avoidance was reported also on other species includ-

ing lobsters, mandrills, termites (Behringer et al., 2006; Poirotte et al., 2017; Rosengaus et al., 1999) and it is generally expected to be favored by selection as it should decrease exposure to contagious pathogens (Barber et al., 2000). However, in some instances, it could be difficult to define whether behaviors after infection are only side effects of an infection, or are driven by an active response (of either the host or the infected conspecifics) (Lopes et al., 2016; Stockmaier et al., 2020,0). For example, immune-challenged bats that reduced social vocalizations also showed social encounters (Stockmaier et al., 2020). Similarly, in case of reduced movement ability due to an infection, we expect self-organizing sorting mechanisms to play an important role in the spatial positioning and possibly the isolation rates of the shoal members (Couzin and Krause, 2003; Couzin et al., 2002). These mechanisms do not invoke complex individual recognition and detection, and may still result in the same outcome with isolation or increased distancing, with similar implications on transmissions.

Finally, another mechanism that can lead to avoidance of unattractive cues and possibly isolate infected individuals is group facilitation and consensus among shoal members (Sumpter et al., 2008). A quorum rule could lead to accurate decisions in avoiding infected group members, as previously shown in sticklebacks making more accurate decisions in avoiding sick-looking conspecifics with increasing group sizes (Sumpter et al., 2008).

To distinguish a passive self-isolation mechanism from an active avoidance mechanism, one possibility is to examine the social interaction structure. While self-isolation should not vary the connectivity among healthy conspecifics, active avoidance should reflect in the strength of the connection of those individuals that actively avoid the possibly contagious one. In our study, we find that infected individuals are more likely to spend time in isolation, as well as to be part of fission from a group size of 8 to a group size of 1. Both these findings are in line with a possible active or passive isolation mechanism. It was previously hypothesized that isolation could occur when the costs associated with parasitic infection outweigh the benefits of being part of a group or when infected individuals are incapable of sustaining shoaling position (Reynolds et al., 2019), for example, due to a lack of motor control in particular for heavily parasitized fish (Krause and Godin, 1996). We previously showed that position in the group is correlated with the increase of isolation rates (Figure B.6), and this trend is observed for both Focal-fish and Non-Focal fish across conditions. This suggests that passive mechanisms such as speed (which also is related to position in the group), play a role in driving the increased isolation of the FF. However, it remains still unclear whether the isolation rates we find are driven by the hosts, or by the uninfected conspecifics.

2.4.3 Infection rates and implications on movement

Unlike infection of parasites with complex life cycles, infections with directly transmitted parasites should also reflect the pressure of the parasite to be successfully transmitted to others, by for example concealing evident behavioral modifications. It was previously shown that guppies in early infections are more attractive than uninfected conspecifics in the same trial (Stephenson et al., 2018), and under semi-natural conditions, infected fish were shown to initiate more social interactions (Croft et al., 2011). It's hypothesized that infected fish might even have generally higher activity levels. In our data, we find similar counter-intuitive observations in a subset of the data, specifically on those fish that show higher relative speed values than other fish in the trial (Figure 2.12). Other evidence of possible effects of concealed early infection signs, is seen in the relationship between proximity and infection transmission. For Non-Focal fish in the trials on Day 5 from the infection (early stage), infection was more likely for those fish that spent a high proportion of time closer than 3 body lengths from the Focal-fish. This result could however, derive from two, not mutually-exclusive hypotheses: in one case, fish at early stages of infection are equally or more attractive than other fish, and thus successfully transmit parasites to those individuals actively spending more time in proximity; alternatively, this could also reflect the 'infectiousness' of the Focal-fish where early stage infections, to be successful, need more time spent in close proximity than later infections do (comparison with data from Day 10, Figure 2.19).

It is known that the main transmission modality of *Gyrodactylus* sp. is by contact with an infected individual (Bakke et al., 2007; Cable et al., 2002). However, it has been also reported that some gyrodactylids might also use a different strategy for transmission that is achieved by detaching from a recently dead individual and moving to the water film (Cable et al., 2002). For the differences we find in the time spent in proximity for individuals that get infected on day 5 and day 10, we hypothesize that the transmission mode might be different. After day 10 of infection, in particular for high infection loads, it is possible that gyrodactylids are also more actively searching for transmission opportunities, and could need less frequent exposure to pass onto the next host. It was previously observed that in some cases the frequency of contacts between individuals, beyond the population density, could drive transmission rates, allowing parasite prevalence also at low densities. Finally, in this study, we do not include information about the Non-focal individual physiological state, in terms of, for example, susceptibility. However, we know from previous literature that susceptibility to parasitic infection, infection development, and transmission probability depend on a combination of physiological, metabolic, and behavioral factors. In future studies, it would be informative to keep information on the progression of the infection on the individuals that get infected during

the trial. This could reveal important information on the susceptibility of the individuals and their behavioral correlates in their pre-infection state (Barber et al., 2017).

2.4.4 Quantifying fission-fusion dynamics: applications and future work areas

In this study, we used a novel approach to quantify guppy's fission-fusion dynamics. This type of animal grouping is highly dynamic both spatially and temporally, where groups split (fission) or merge (fusion) at different timescales varying group sizes and composition. Here, we used a distance threshold to segment trajectories, define group membership, and classify events as a subset of coordinates of consecutive frames where group composition (based on fish identities) is fixed. With this method we add extra detail to the quantification of the collective behavior of the group, quantifying the timing of different group sizes, group composition, and transitions across groups. This level of detail is not present in, for example, studies with golden shiners (Couzin et al., 2011; Tunstrøm et al., 2013) or locusts (Buhl et al., 2006), or other species where the group generally stays together. With this approach we can quantify the level at which each individual group member contributes to the process, and how parasitism affects these dynamics. Other animal groups such as ungulates, elephants, cetaceans, or baboons also display fission-fusion behavior (Couzin and Laidre, 2009). While we applied and developed our approach with guppies, the approach is general, and future work can build these methods to quantify the fission-fusion dynamics of other collectives.

2.4.5 Conclusion

In this chapter, we described the range of variation of the behavior of shoaling guppies and how the individual and collective behavior varies in presence of a parasitized fish. Overall, we find evidence for changes in the behavior in response to parasite infection at the individual-level, which could generate cascading effects at the group and population levels. We find that the presence of the infected fish affects group cohesion - increasing nearest-neighbor distance and increasing the likelihood for the infected fish to fission to isolation. We also find a re-organization in the group positioning, where infected individuals are more likely to take peripheral positions (and more in the front), which could be a mechanism to minimize phenotypic heterogeneity while decreasing the infection risks. Overall, we think that the infection of *Gyrodactylus* sp. does not induce uni-directional effects on the behavior of the group, but rather increases the likelihood of some behaviors appearing, at a higher magnitude than of non-infected fish.

Chapter 3

Analysis of leader-follower relationships of shoaling guppies across timescales

3.1 Introduction

Interactions between individuals play a critical role in shaping animal behavior. The specific rules that animals use to interact with each other can have a profound effect on the overall behavior of the group, and such effects can have cascading effects on how individuals in groups ultimately make collective decisions (Couzin and Krause, 2003). In some cases, decision-making involves a degree of leadership. In a broad definition, leadership is a process of differential influence initiated by one or more individuals that results in behavioral change of other group members (Krause et al., 2000; Strandburg-Peshkin et al., 2018). Members of pigeons' flock exhibit a hierarchical structure in terms of who influences whom in terms of group motion, as revealed by characteristic delays in time between individuals' directional choices (Nagy et al., 2010). By contrast, individuals' influence on collective motion can be more egalitarian, where leader-follower relationships are not rigid and individual roles vary over the course of the interaction (Dyer et al., 2009). Thus there exists a continuum between diffuse and highly-hierarchical leadership in coordination movement of animal groups. In this chapter, we explore an analysis of leader-follower dynamics of fish shoals. However, unlike previous studies of differential social influence, which rely on instantaneous, thus short timescale velocity vector information, as previously done in Nagy et al. (2010), we expand the analysis to also describe leader-follower dynamics of activity level changes at long timescales.

The behavior of animals is multifaceted and it may vary widely over their lifetime. Similarly, patterns of associations in animal groups also change across timescales, from

near-instantaneous fusion in response to increased risk of predation to daily fission-fusion patterns, to longer timescale seasonal migration. While factors such as sex are (mostly) stable over the lifetime of an animal, aspects like size, locomotion, pathogen infection, or nutritional state, may vary over shorter timescales (Jolles et al., 2020a). There has, however, been surprisingly little quantitative work to account for variation of individual behavior across timescales, likely due to limited resources and technology for lab and field behavioral experiments. In this chapter, we explore data from 55 recordings of eight freely shoaling guppies. For each group, we analyze two hours of data and describe how individual-level speed/activity patterns vary, within the duration of the experiment, at short and long timescales, and how the degree of coordination with the other shoal members shapes leader-follower dynamics across timescales.

Both modeling and empirical data have investigated the mechanisms of leadership and decision-making of moving animal groups, showing that leadership and signaling among individuals are not necessarily driven by dominance hierarchies (Couzin et al., 2005; Reader et al., 2003; Reebs, 2001,0; Strandburg-Peshkin et al., 2015). Work on fish schools has shown that a fraction of individuals can lead larger groups of conspecifics to a targeted area (Couzin et al., 2011; Reebs, 2000), that body size or nutritional state influences position in the group and eventually group movements (Hoare et al., 1998; Reebs, 2001), and that quorum decision-making is used to inform decision making and information transfer (Ward et al., 2008). A common example of how animals within groups must conform with others to maintain group cohesion is by adjusting their speeds (Herbert-Read et al., 2013; Jolles et al., 2017; Sankey et al., 2019). In this study, we look at how small fish groups coordinate their activity, social cohesion, and social coordination by adjusting their speed, and how this level of organization varies across timescales of interactions. To this end, we analyzed the time-shifted correlation of speed to infer leader-follower dynamics of guppy shoals, and how these vary across a broad range of timescales. We find consistent inter-group differences and we describe correlation across activation levels of different individuals of the same population.

To investigate the influence that fish have on the other shoal members, we evaluated the temporal relationship of speed for each dyad in the group. We analyzed individuals' motion at different timescales by considering individuals' swim speed transformed with different moving average window sizes, as will be detailed below. To describe short timescale changes in speed, we consider a small moving window size, such that this represents instantaneous changes, and we refer to this as the animal's speed. To describe long timescale changes, we use a larger window, and because of the larger window size, this can be better thought of as representing the 'activity level' of individuals. For the purpose of this chapter, we focus most of the analysis on comparing differential social

influence (*i.e.*, who most influences whom - a form of leadership) calculated at one short timescale (average over 1 second) and one long timescale (average over 150s). However, where needed we also include leadership scores calculated on additional timescale values. For each dyad in each experiment, we used a method based on cross-correlation of speed (and activity) time series, similar to that previously developed by Nagy et al. (2010). This allows us to infer a leader-follower structure that informs us about the 'initiators' and 'followers' of the group regarding both short and long timescale changes in speed and activity, respectively. While previous work has identified factors that predict short timescale leadership, we use our data to ask if the same hierarchies and predictors apply at both short and long time timescales. Finally, we discuss preliminary data on how pairwise temporal dynamics across timescales also vary as a function of the parasite load of group members.

3.2 Methods

3.2.1 Dataset

In this study, we included all data used for those studies described in Chapter 2. These data contain trajectories of all individuals in freely-swimming shoaling guppies *Poecilia reticulata*. Groups of eight guppies were filmed swimming in a 116x116cm experimental arena at 30 Hz for three hours. The videos were analyzed using TRex (Walter and Couzin, 2021), software that both allows individual tracking and individual identification. We analyzed a total of 55 recordings, each of which is referred to as one trial. The data from these trials were originally used for an analysis of the effect of parasitism on shoaling, where the infection status of one individual of the shoal was experimentally manipulated. More details on the infection procedure and the details on the experimental trials are found in sections 2.2.2 and 2.2.3 of Chapter 2.

3.2.2 Capturing leadership dynamics across timescales

From the analysis of the speed time series, one for each individual, we excluded the first hour of the experiment after an evaluation of the acclimation time that occurs on average on all trials; thus each trial is 2 hours in duration. To read more about the experiment procedure, see section 2.2.4 and Figure 2.5 of Chapter 2, where we describe in detail how tracking was conducted and trajectories obtained using TRex (Walter and Couzin, 2021). After initial filtering of the trajectories to exclude tracking errors, we obtain the position of each fish i at time t and we calculate their instantaneous speed $u_i(t)$ as the norm of the velocity vector $v_i(t)$. From the instantaneous speed value $u_i(t)$ we then used a moving

average with different window sizes w to transform the data. The window sizes (*i.e.* timescales) are – $w \in$ {0.26 s, 1 s, 5 s, 15 s, 30 s, 60 s, 150 s, 300 s} (Table 3.1). Shorter window sizes represent (nearly) instantaneous speed, while longer window sizes are a representation of activity. Preliminary analysis suggested a continuum in the values calculated at each window size w; therefore, to simplify the quantification, we use the window w of 1 second as the speed at short timescale, and the window w of 150 seconds as the activity at the long timescale. More specifically, we selected a moving average of a short window size w of 1 second (30 frames) since this still captures variation in the individual speed at a temporal resolution associated with the timescale of fine-scale movement dynamics within these groups. In other words, this value represents short-timescale changes of speed, representing the relatively rapid speed and direction changes exhibited by guppy groups. By contrast, applying a moving average on the speed with a relatively large window size of 150 seconds (4500 frames), results in short timescale information (*e.g.* tail-beat frequency, or burst and glide) being hidden, while revealing a long timescale representation of the speed, which can better be thought of as overall 'activity level'. In other words, what the activity level represents is a quantification of how generally more or less active fish are, and how such long-timescale changes in activity vary over the course of the trial. While the main results compare the short and the long timescale, for some analysis we also include the leadership scores calculated at each window size w of the range described above.

Figure 3.1 shows an example of the averaging computation of the speed time series. In the top part of Figure 3.1 we see the data for 8 individuals with individuals' speed averaged with a window size of 150 seconds, thus representing the 'activity level' for each, shown for the entire duration of the trial (three hours). What we notice is a general correlation across the individuals, with fluctuations evident in this relatively long-timescale level of activity. The bottom part of Figure 3.1 shows a subset of the speed data averaged using different window sizes, looking at one time segment of the trial. The subset here is around 55 minutes in duration and window sizes of 1 second, 30 seconds, and 150 seconds. The time series obtained with a moving average of 1 second is similar to the instantaneous speed and thus highly informative regarding short timescale dynamics. By applying moving averages of increasing window length, we obtain information that loses the short timescale information (*e.g.* tail-beat), but describes a more general system activity level. For example, the orange time series obtained with a window size of 30 seconds still shows how individuals move in bursts, on top of a more general activity level that is better captured by the longer window size, here displayed in purple, with a moving average window of size 150 seconds (Figure 3.1).

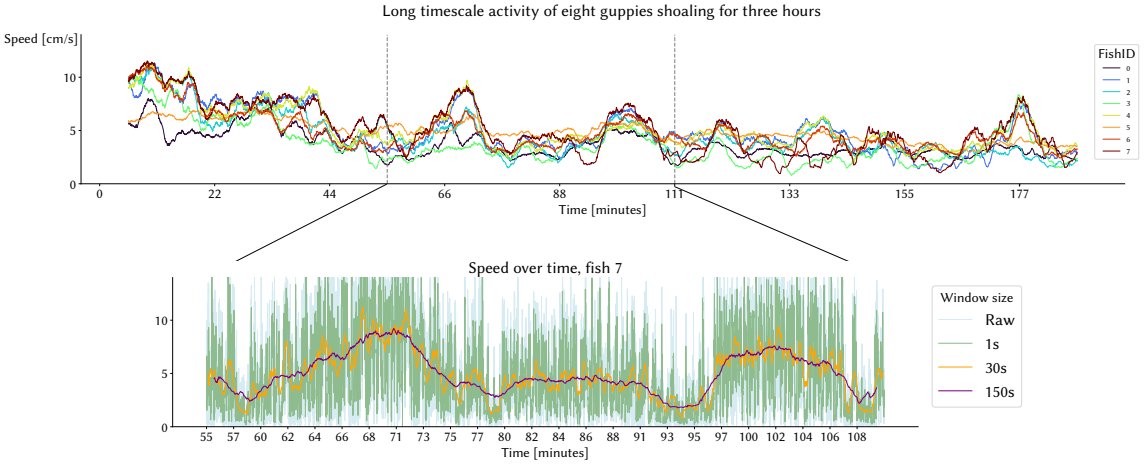


Figure 3.1: **Speed values at timescales.** From the raw speed value of each fish, we used a moving average to analyze the data across eight different window lengths (*i.e.* timescales) – $w \in \{0.26$ s, 1 s, 5 s, 15 s, 30 s, 60 s, 150 s, 300 s$\}$. Top: long timescale activity level time series of eight individuals in the same trial, obtained with a window size of 150 seconds. We visualize the activity traces for the entire duration of the trial (three hours). Bottom: comparison of speed/activity time series across different timescales. Raw and short timescale speed time series still contain high frequency signals, with higher resolution at the short scale, thus possibly informative on short range locomotion and movement dynamics. Longer window sizes of 30 and 150 seconds lose short timescale resolution (*e.g.* tail-beat frequency), but reveal what we here refer to as 'activity level', where fish are described as generally faster or slower (more or less active).

3.2.3 Speed/activity correlation and leadership calculation

To quantify the reciprocal social influence among all individuals in a trial, we used a cross-correlation analysis of the speed time series averaged at either 1 second, to capture fast-timescale influence, and 150 seconds, to capture how individuals influence each others' overall activity levels throughout the trials. This method is based on Nagy et al. (2010).

Consider the speed time series of two individuals i and j, which have been averaged with a window of w, and denote these as $s_i(w,t)$ and $s_j(w,t)$. Dropping the w for ease of notation, the correlation between these individuals for a time lag of τ is calculated by averaging over each trial:

$$C_{ij}(w,\tau) = \frac{\int dt\,(s_i(t) - \bar{s_i})\,(s_j(t-\tau) - \bar{s_j})}{\sqrt{\int dt\,(s_i(t) - \bar{s_i})^2 \int dt\,(s_j(t-\tau) - \bar{s_j})^2}} \tag{3.1}$$

where $\bar{s_i}$ and $\bar{s_j}$ denotes averages. For positive τ values the integral goes from $[\tau, T]$, and for negative τ values from $[0, T+\tau]$, where T is the total trial duration for analysis.

To calculate C_{ij} numerically, we calculated $C_{ij}w,\tau$ over a range of $-A_w$ to A_w, with a step size of Δ_w. The numerical values of A_w and Δ_w are listed in Table 3.1.

For each fish pair within a trial, we obtained the time lag Γ_{ij}^w, which is the value τ at

the maximum value of the correlation function $C_{ij}(w,\tau)$. Positive Γ_{ij}^{w} values correspond to individual fish i that initiate speed/activity changes prior to the individual fish j, whereas negative Γ_{ij}^{w} are found when the individual fish i 'follows' the speed/activity changes of the individual fish j. For an illustration of this calculation see the Appendix C.1 and Figure C.1.

Therefore, for each fish pair we have

$$M_{ij}^{w} = \begin{bmatrix} 0 & M_{12} & M_{13} \\ M_{21} & 0 & M_{23} \\ M_{31} & M_{32} & 0 \end{bmatrix} \tag{3.2}$$

which is symmetric and

$$\Gamma_{w,ij}^{w} = \begin{bmatrix} 0 & \Gamma_{12} & \Gamma_{13} \\ \Gamma_{21} & 0 & \Gamma_{23} \\ \Gamma_{31} & \Gamma_{32} & 0 \end{bmatrix} \tag{3.3}$$

which for each fish pair is anti-symmetric.

To ensure selection of pairs with correlation values that hold meaningful information about speed changes, we selected pairs with a $M_{ij}^{w} > M_{tr}^{w}$, depending on the window size w. We also disregard events where the time lag $\Gamma_{ij}(w,\tau) < \Gamma_{max}(w,\tau)$, as this time lag would be too short for the influence of speed change to be meaningful. In order to use the data to set a correlation threshold value, the value of M_{tr}^{w} for each window size w was decided by the maximum $C_{ij}(w,\tau)$ over the entire window $[-A_w, A_w]$ among 100 random pairs $s_i(w,t)$ and $s_j(w,t)$ of fish i,j selected from different trials, therefore fish that were not in a tank at the same time. This was done in order to estimate a meaningful threshold for the correlation and can be considered to be representative of what we expect as values for random correlations among fish (Figure C.2). Finally, M_{tr}^{w} was chosen as the 95^{th} percentile of the maximum $C_{ij}(w,\tau)$ distribution for each window size w.

For each fish pair in a trial, using the maximum correlation M_{ij}^{w} and the time lag at the maximum correlation Γ_{ij}^{w}, we then created a matrix that represents whether or not the pair had a meaningful leader-follower relationship over the course of the trial:

$$S_{ij}^{w} = \begin{cases} 1, & \text{if } (|\Gamma_{ij}^{w}| > \Gamma_{min}(w,\tau))\&(|\Gamma_{ij}^{w}| < \Gamma_{max}(w,\tau))\&(M_{ij}^{w} > M_{tr}^{w}) \\ 0, & \text{otherwise.} \end{cases} \tag{3.4}$$

Using the matrix S_{ij}^{w}, which thresholds for meaningful leader-follower relationships, we defined a scaled individual leadership score as

$$L_i^w = \frac{1}{n}\sum_{1}^{n} \text{sgn}(\Gamma_{ij}^w) \cdot M_{ij}^w \cdot S_{ij}^w$$

Note that the scaled relative leadership score is normalized by the maximum possible score, which corresponds to an individual that leads (positive score), or follows (negative score) all others, with the maximum correlation value being 1.

An overview of the window sizes (w) used, and the respective parameters and numerical values are found in table 3.1.

Timescale [s]	Equation parameters [frames]			Numerical parameters [frames]	
w	Γ_{min}^w	Γ_{max}^w	M_{tr}^w	$[-A_w$ to $A_w]$	Δ_w
0.26	1	60	0.15	± 70	1
1	2	130	0.15	± 150	2
5	2	160	0.2	± 180	2
15	5	420	0.25	± 450	5
30	10	800	0.3	± 900	10
60	20	1500	0.3	± 1800	20
150	50	3500	0.45	± 4500	50
300	50	4000	0.6	± 5000	50

Table 3.1: **Parameters leadership score.** We compared a total of eight timescales w of length ranging from 0.26s to 300s. For each w we defined parameters of Γ_{min}^w, Γ_{max}^w used to disregard events where the time lag $\Gamma_{ij}(w, \tau)$ would be too short for the speed change to be meaningful in being copied. The value of M_{tr}^w for each window size w was decided by calculating $C_{ij}(w, \tau)$ among 100 random pairs $s_i(w, t)$ and $s_j(w, t)$ of fish i, j selected from different trials, therefore not in a tank at the same time, and used to ensure selection of pairs with correlation values that hold meaningful information about speed changes. The values of the numerical parameters for the range $[-A_w$ to $A_w]$ and the step size Δ_w are shown in the last two columns of the table.

3.3 Results

For each pair of individuals in a trial, we calculated a leadership score to evaluate the relative influence an individual i has on the group in determining speed/activity changes. In total, we computed leadership scores (L_i^w) for eight timescales (Table 3.1). To examine core trends of leadership at different timescales, we focus our analysis to compare a short timescale (w = 1s) and a long timescale (w = 150s).

3.3.1 Leadership scores across trials

Leadership scores are calculated for each fish, in each trial. Figure 3.2 shows all leadership score values for each trial. Positive leadership scores indicate individuals that

initiate a speed/activity change that is subsequently exhibited by other group members. Negative leadership scores present individuals that tend to copy the speed change initiated by other group members. Leadership scores close to zero indicate either individuals with non-rigid leadership roles or individuals that are uncorrelated to the rest of the shoal and thus, do not exhibit a relationship with other shoal members. The magnitude of the leadership score is indicative of the strength of the correlation with other shoal members, as well as the consistency of the leader (positive) or the follower (negative) role. In Figure 3.2 we visualize the variation in individuals' leadership scores for each trial (each box plot), in both the short (Figure 3.2 top row) and the long timescale (Figure 3.2 bottom row). We note that that the within-trial variance of leadership scores is similar for the two timescales analyzed. This suggests that for trials where there exists a leadership structure over a short timescale there may also be one over the longer timescale.

The within-trial distribution of the leadership score describes the structure of leader-follower dynamics, from which we infer the hierarchy determined by the ranking and magnitude of leadership scores of each fish in the trial. Overall, we find that some trials are characterized by a small variance, both at short and long timescales (left-most trials in Figure 3.2), which reflects a flat structure, where no individual consistently leads initiations of changes in the speed dynamics (or do not interact meaningfully, thus low correlation with other individuals), while others show clear leader and follower individuals (right-most trials in Figure 3.2), with higher variance in leadership scores.

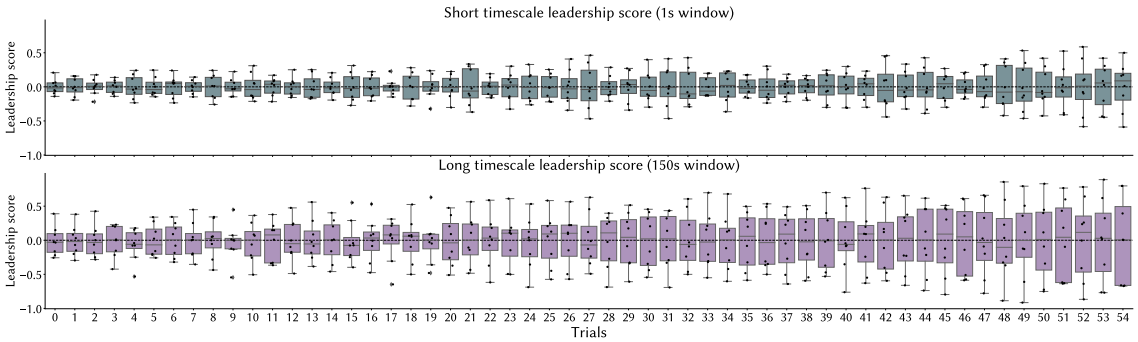


Figure 3.2: **Distribution of leadership scores for each trial at short and long timescales.** For each trial we show the distributions (boxplots) and the leadership scores obtained for each individual in the trial (dots). Values are sorted in increasing order by leadership score of the long timescale (bottom row). Positive leadership scores indicate individuals that initiate a speed change that is followed by other group members. Negative leadership scores are found for individuals that copy the speed change initiated by other group members. Box plots give the median, interquartile range (box), and values that sit outside of 1.5 times the interquartile range (whiskers).

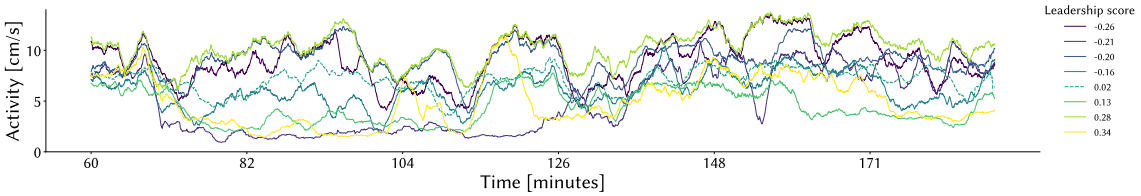
3.3.2 Illustrative example: comparison of activity levels in two trials with different leadership score distributions

To better understand what leadership at long timescale activity levels can represent, and how trials differ, we looked at two example trials that differ by the variance of the distribution of the leadership scores. In Figure 3.3, we illustrate the long timescale activity levels within a trial characterized by a small variance in leadership scores (variance = 0.05, Figure 3.3a), and compare it to a trial in which there was a large variance in the leadership score at the long timescale (variance = 0.39, Figure 3.3b). The activity time series of each individual in Figure 3.3 are colored by the individuals' leadership score. In both examples, the activity level time series of each individual varies over the course of two hours, alternating phases of high and low activity levels in the order of 5 or 10 minutes, to phases of activity that last up to 40-50 consecutive minutes. As an example, note the activity level of one of the individuals in Figure 3.3a (dark blue line, leadership score = -0.21), with a constant low activity level between minutes 70 and 125. Overall, the low variance in leadership scores in Figure 3.3a indicates the absence of consistent leaders and followers, as evidenced by generally low synchrony of the activity changes among group members, and differences in the absolute swim speed (activity). This is different from what we observe in Figure 3.3b, where individual activity time series are largely overlapping and show similar increasing and decreasing activity values over time. The large within-trial of the leadership score of the trial in Figure 3.3b suggests that the collective dynamics are synchronized and that there exist consistent leaders and followers.

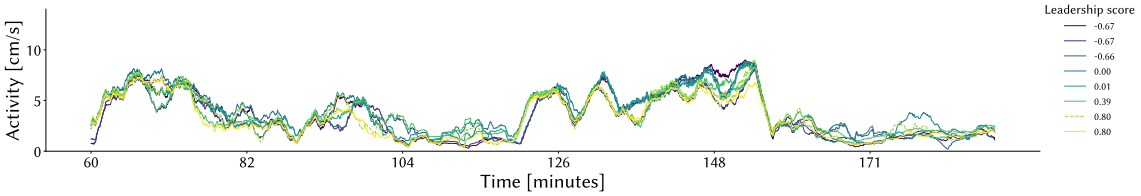
In this section, we used examples from our data to show how the analysis method on long timescale time series describes structures in the group movement patterns. The examples taken, are visually different, describing variations on both individual and collective movement patterns. The first example showed a trial where speed time series are uncorrelated, and where there is no clear leader-follower structure (Figure 3.3a). The second example, described a trial with high variance in the leadership scores, indicating a clear structure in the leaders-follower roles of the group, which in this case also represented by high group cohesion, and low speed difference among the shoal members (Figure 3.3b).

3.3.3 How individual scores vary across timescales

We next asked if fish that exhibit leadership over short timescales are also leaders with respect to long timescale (*i.e.*, activity level) dynamics. To do so, we asked if individuals' leadership scores for the short timescale are correlated with their leadership scores over the long timescale. For this analysis, we include data on leadership scores calculated



(a) Long timescale activity time series - trial with low leadership score variance



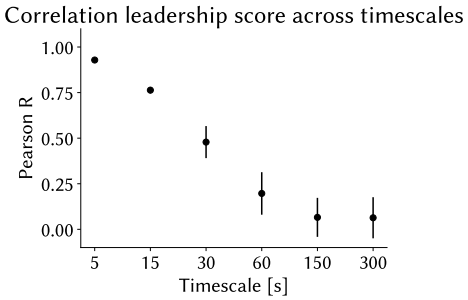
(b) Long timescale activity time series - trial with high leadership score variance

Figure 3.3: **Long timescale activity time series in two trials.** Here, we compare two trials that differ in the variance of leadership scores. We therefore compare the activity time series of the long timescale. (a) Example trial with weak leader-follower structure (small leader-follower scores and variance). The leadership scores of the shoal members of this trial are shown in Figure 3.2, trial 6. (b) Example trial with high pairwise correlation and clear leader-follower dynamics. The leadership scores of the shoal members are shown in Figure 3.2, trial 54. In this second example, we also notice low speed difference among the shoal members which in this case denotes high group cohesion.

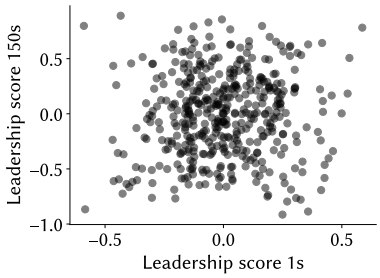
at each of the eight window sizes w (Table 3.1). We find that the correlation decreases with the increase in the time lag between the timescales tested (Figure 3.4a). The leadership score calculated in the short timescale of 1s highly correlates with the following timescale of 5s (r=0.84, $p < 0.001$), but progressively decreases with increasing timescales (Figure 3.4a). Leadership scores are correlated in pairs of timescales of increasing sizes, suggesting that information at each timescale is progressively changing but overlapping (Figure 3.4c); however, the correlation between leadership scores of the short (1 second) and the long (150 seconds) timescales is close to zero (r=0.06, $p > 0.05$) (Figure 3.4b). These results suggest that the leadership scores calculated over a short timescale are not predictive of the leadership scores calculated at a longer timescale, suggesting that different individuals are more, or less, influential over different timescales (*i.e.* those who lead short-timescale schooling dynamics are not those who influence the overall activity level exhibited by individuals in the population).

3.3.4 Comparison of leadership scores and behavioral metrics

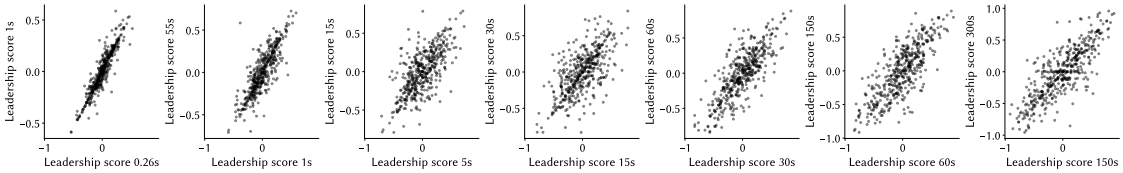
In this section we investigate the correlation between individual/group-level behavioral metrics and leadership score at the short timescale of 1s, and at the long timescale of 150s (Figure 3.5). All results in this section report the Pearson Correlation Coefficient and the significance p-value. The metrics with the strongest correlation to leadership



(a) Correlation of the leadership score across timescales



(b) Correlation leadership scores at short and long timescales



(c) Correlation of neighboring timescales

Figure 3.4: **Correlation leadership scores at different timescales** (a) Correlation of the leadership scores of the short timescale, with leadership scores of increasingly longer timescales. The correlation decreases with the increase in the time lag between the timescales tested. The error-bar shows the variance across individual trials. (b) The correlation between the leadership scores calculated at short timescale (1s window) and at long timescale (150s window) is close to zero (r=0.06, p > 0.05). (c) Leadership scores correlate between neighboring timescales (in pairs),

scores in the short timescale are front-back (FB) distance and relative acceleration (r=0.6, p < 0.001) (Figure 3.5, top row). These results confirm previous studies, where leader-follower dynamics were evaluated at short timescales (Jolles et al., 2017; Nagy et al., 2010; Sridhar, 2022). Relative speed, turning rate, and speed variance (r=0.3, p < 0.001), show a weak correlation with leadership scores at the short timescale, followed by nearest-neighbor distance (r=-0.2, p < 0.001) and time spent in isolation (r=-0.1, p < 0.05) (Figure 3.5, top row). The same seven metrics, however, do not exhibit the same relationship with the long timescale leadership score, either showing a low or near zero positive correlation (turning rate, relative acceleration, speed variance, nearest-neighbor distance) or a weak negative correlation (front-back distance, relative speed) (Figure 3.5, bottom row). This suggests that mechanisms of leadership at long timescales are different than the mechanisms of leadership at short timescales.

3.3.5 Leader-follower dynamics and the effect of parasitism

Here we explore how leader-follower dynamics relate to parasite load of the infected individual in each trial. Previously we have found that parasitism can affect short-time

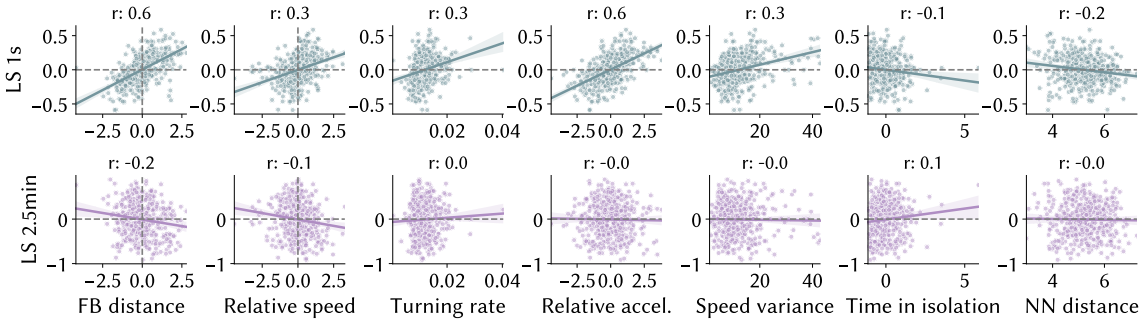


Figure 3.5: **Correlation of individual and group behavioral metrics and leadership scores.** Correlation between individual and group metrics to the leadership score at the short timescale of 1s (top row), and at the long timescale of 150s (bottom row). The metrics with the strongest correlation in the short timescale are front-back (FB) distance and relative acceleration (r=0.6, p < 0.001), followed by relative speed, turning rate, and speed variance (Pearson Correlation Coefficient, r=0.3, p < 0.001). None of the tested metrics shows correlation with the long timescale leadership score (bottom row). All r coefficients are calculated with the Pearson Correlation Coefficient.

scale dynamics: for example, parasitized individuals show higher nearest-neighbor distance than non-parasitized fish and also occupy more frontal and peripheral positions in shoals of different sizes (see Chapter 2). In this analysis, we only include data for fish measured before and after infection. More specifically, we compare data of leadership score of the Focal-fish at the Baseline and at the Treatment condition, which details are explained in Chapter 2, section 2.2.3.

Regarding social influence both at the short and long timescale (Figure 3.6, left and right, respectively), we observe that for increasing parasite load, the magnitude of the leadership score decreases (Figure 3.6). At the short timescale, for fish with no or low parasite loads, we observe a larger variance with leadership scores ranging from -0.22 and 0.49 (minimum and maximum values of the distribution), compared to leadership scores for individuals with higher parasite loads at the same timescale. At the long timescale of 150s, we also find a similar relationship between parasite load and leadership score, with decreasingly lower leadership scores at increasing parasite load. Moreover, at the long timescale, the leadership score values are generally higher in magnitude and range between -0.90 and 0.92 compared to the short timescale. These results suggest that parasitism may have an effect on the role of the infected fish, by modifying its interaction with the other group members, and that the relationship is similar at the short and the long timescale.

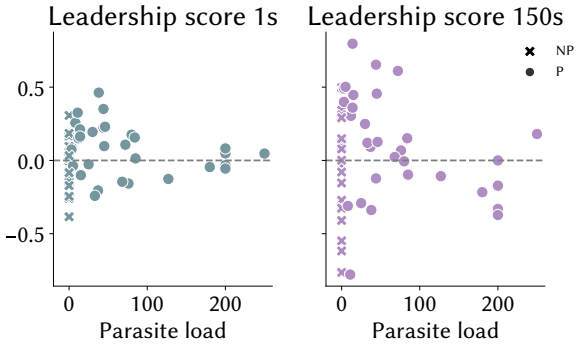


Figure 3.6: **Leader-follower dynamics as a function of parasite load** (a) Correlation between parasite load and leadership score at the short timescale of 1 second. Leadership scores are generally smaller at the short timescale than they are at the long timescale (b). However, we still obtain information on the direction of some of the meaningful pairwise interactions. Overall, we find a larger variance in the leadership scores of individuals with small parasite loads, as opposed to individuals with high parasite loads. (b) Correlation between parasite load and leadership score at the long timescale of 150 seconds. At the long timescale, we find that the variance of the leadership scores is generally higher than that of leadership scores at the short timescale. Low parasite loads show a relative increase in the leadership scores compared to non-parasitized fish (marked with 'x'), followed by a decrease in the scores for increasing parasite loads (marked with 'o'). Non-parasitized fish are marked with the symbol 'x', to distinguish from parasitized fish marked with the symbol 'o'.

3.4 Discussion

In this chapter, we explored how individual and collective values of speed/activity change over two hours of movement data of freely shoaling fish, and how members of a shoal coordinate and adjust their speed to the other shoal members over different timescales. The individual values of the speed of each shoal member were transformed at different timescales by applying a moving average of different window sizes to the raw speed data, thus describing coordination at different temporal resolutions, where the short timescale window represents 'speed', while the long timescale window represents 'activity level'. For each of the eight timescales tested, we calculated leader-follower dynamics based on a pairwise correlation analysis between shoal members. This procedure is based on methods developed by Nagy et al. (2010) which evaluates the instantaneous temporal relationship of pigeons' directional correlation. Instead of direction correlations, we use only speed values to analyze leader-follower dynamics; this simplification enables us to examine different timescales of activity levels and the resulting differences in leader-follower relationships. We find that individual leadership scores for the short timescale are not predictive of who is most influential over a long timescale (Figure 3.4b). Behavioral metrics that are commonly used to describe instantaneous kinematics positively correlate with leadership at the short timescale, but this correlation is lost over a longer timescale (Figure 3.5). Overall, this suggests that the mechanisms that govern leadership

at the short timescale are likely different from those over the long timescale, and that common behavioral metrics might be insufficient to describe coordination and patterns over multiple temporal scales.

Trials were found to vary with respect to the observed leader-follower structure. In some trials, all members exhibited low leadership scores, whereas in others we found a clear distinction between, and consistency within leaders (high positive leadership scores) and followers (low negative leadership scores) (Figure 3.2). By selecting two trials with characteristic low and high variance of leadership scores, we illustrated how long timescale activity time series, as shown in Figure 3.3 reveal patterns of variation over time for each individual. This suggests that longer behavioral analyses, such as over a few hours, as opposed to the more typical timescale of minutes, can improve the evaluation of inter-individual correlation, group cohesion, and leader-follower relationships.

Groups of animals consist of heterogeneous individuals, and in recent years there has been increasing interest in understanding how collective behavior patterns result from individual-level processes. In the context of 'consensus' and decision-making in animal groups, it is not unusual to find differences in how individuals in groups coordinate, and vary inter-individual distances according to need. Animals constantly need to adjust behavior in response to others, and the environment, and 'consensus' among members of both animals and humans is thought to ensure coherent collective action, despite differences in individual preference or physiological requirements (Conradt et al., 2009; Dyer et al., 2009; Herbert-Read et al., 2019; Sumpter and Pratt, 2009). For example, the decision-making process to move towards a common target, or to synchronize a group activity, can involve group coordination and consensus, and often consensus costs are lower for a group, as opposed to the outcome of unshared decisions and thus, possibly leading to fragmentation (Conradt et al., 2009; Conradt and Roper, 2003). The variation we observe across trials is likely indicative of individual differences, and a better understanding of group coordination and leader-follower dynamics at different timescales could reveal important information regarding the role of individuals within the group, and the resulting group behaviors exhibited.

Previous studies have shown that instantaneous kinematics measures are correlated with leadership scores over the same short timescale (Jolles et al., 2017; Nagy et al., 2010; Sridhar, 2022), and in particular that faster individuals that occupy frontal positions increasing their propensity to lead others (Rosenthal et al., 2015). With this study, we confirm such findings for short timescales. We find that the leadership scores calculated at the short timescale are positively and strongly correlated with metrics of front-back spatial positioning in the group, as well as with mean acceleration calculated relative

to the acceleration of shoal members. Other values of relative speed, turning rate, and speed variance also show a positive correlation with leadership, although with a weaker correlation (Figure 3.5). The correlation of front-back spatial positioning, acceleration, and speed metrics with leadership scores at the short timescale is expected, as the individual leadership scores are calculated on the basis of pairwise interactions, and therefore describe the role of an individual to either 'initiate' or 'follow' the speed changes of other individuals. Theory and empirical data show that if individuals move together, they will naturally adjust their motion in order to maintain proximity. By contrast, it is possible that the activity level at the longer temporal scale is less dependent on spatial proximity, meaning that the adjustments of activity levels to match those of the group members are less predicted by metrics that we commonly use to describe instantaneous spatio-temporal dynamics. In this study, when we consider longer timescales, we find that the leadership scores do not correlate with measures of short-range spatial interaction or short timescale movement metrics (*e.g.* spatial position in the group, nearest-neighbor distance, relative speed) (Figure 3.5).

Group spatio-temporal dynamics also have important implications on disease dynamics, where increased associations in smaller communities might increase the transmission risk, in particular when a parasite is transmitted via direct contact (Barber et al., 2000). Across taxa, we find different mechanisms used to respond to disease threat, but it is challenging to understand how within-individual variation scales to population-level structures, and how this might affect dynamics in the presence of disease (Stockmaier et al., 2021). Pathogen infection is an example process that has implications over a range of timescales, from the short-term modification of for example locomotion patterns, to processes such as isolation and avoidance, that might happen over minutes, hours, or days after initial exposure (Jolles et al., 2020a). Therefore, in this chapter, we also evaluated how the leader-follower role of a diseased individual varies as a function of parasite load, across timescales. We find that the magnitude of the leadership scores decreases for increasing parasite loads. Low values of leadership scores indicate either cases in which the leader-follower dynamics vary throughout the trial, which results in an average of zero, or a low overall number of interactions with shoal members. To test if the decrease in leadership score is indeed due to lower rate of interactions, we can compare the (non-time-lagged) correlation between fish. This measure is indicative of the correlation of the speed/activity time series between pairs of individuals, independent of the leader-follower relationship. Figure C.3) shows a decrease in correlation among shoal members at increasing parasite loads (Figure 3.6). Therefore, this suggests that in this study, the decrease in the magnitude of the leadership score of the infected fish is likely driven by an overall decrease in the correlation of the speed and activity patterns of the

infected fish, with the rest of the group.

To conclude, in this chapter, we explore how individual-level speed patterns vary across short and long time scales, and how these describe the degree of coordination within fish shoals. By calculating pairwise cross-correlation values across speed and activity levels with time delay, we calculated leadership scores for each fish, and we compared the results across different timescales. This study is exploratory and looks at longer timescales of motion, and we note some limitations of the current approach. One limitation is the difficulty in interpreting the role of individuals that have leadership scores close to zero. As previously observed, the absence of strong leadership is currently possibly the outcome of two different mechanisms. On the one hand, it could be the result of a lack of time-lagged correlation with other individuals (below the threshold), and therefore would not been taken into consideration for the calculation of the leadership score. On the other hand, a leadership score of zero could reflect a group structure where leader-follower continuously varies over the course of a trial. Finally, the current approach does not take into account the direction of the activity change, which we normally describe as a general activation or deactivation of the group. In the future, it will be informative to differentiate between individuals that trigger a meaningful increase or decrease of the activity. We hypothesize that this information could be informative also in the context of parasitism and parasite transmission, where infected individuals could be distinguished from uninfected individuals for the role in either initiating increases or decreases in the group activity level.

General Discussion

In recent years, scientists have increasingly employed machine learning tools, such as deep neural networks, to measure animal behavior. This approach allows for the direct recording of details of animal behavior from real data efficiently and without the need for marking or tagging animals. With these tools we can gain behavioral insights from an animal's posture – the time-varying positions of an animal's body parts in space – as well as from the animal's position in space over time. Ultimately, these techniques facilitate a new understanding of animal behavior and can provide new insights about how the brain organizes behavior. Here, we present how we used a quantitative approach and applied modern image-analysis software technology (Mathis et al., 2018; Walter and Couzin, 2021) to better understand the role of disease in a social species.

With this work, we looked at the effect of a socially-transmitted parasite *Gyrodactylus* sp. on the behavior of guppies, from the individual to the collective level. At the individual level, we combined posture estimation analysis and respirometry to characterize the swimming behavior of infected guppies and compared it to uninfected conspecifics (Mathis et al., 2018). At the collective level, we used cutting-edge tracking technology (Walter and Couzin, 2021) to extract highly detailed information on the movement of shoals of guppies, to compare 'healthy' shoals to shoals with a parasitized individual. Finally, we explored differential social influence (normally referred to as leader-follower dynamics), in guppies shoals at different timescales, both in the presence and absence of contagious parasites.

In Chapter 1, we compared the swimming behavior of parasitized and non-parasitized guppies using measures of physiology (oxygen consumption rates), kinematics, and swimming performance (maximum critical speed) across increasing flow speeds. We found that parasitism has a complex effect on locomotion, with parasitized fish showing both increased and decreased swimming speeds, depending on the parasite load. This suggests that the effect of ectoparasites on locomotion is not uni-directional, but the quantitative approach we developed describes the range of variation of swimming behavior in relation to parasitism, locomotion, morphology, and physiology. Future studies should expand on the role of individual traits in determining the large variation

in swimming styles we observed in the population, for example by looking at individual susceptibility or resistance measures. Overall, even in the absence of discrete swimming mode categories, by using a combination of metrics we find within-species variation that informs us regarding inter-individual phenotypic variation that can be linked to different behavioral types. Understanding how locomotion and physiology are impacted by the pathogen can inform us on the mechanisms that induce behavioral modification of an infected individual, and on the possible cascading effects these may have in the social context.

An individual's behavior can impact pathogen transmission, and pathogens can spread quickly through a population, causing serious illness or even death. In the case of contagious diseases, pathogen transmission can be affected by how often individuals come into contact with each other and how long they remain in contact (Stockmaier et al., 2021). In order to better understand the role of parasitism in the social context, in Chapter 2 of this thesis, we used tracking technology to analyze the individual motion features of freely shoaling guppies in the presence of a socially transmitted parasite, *Gyrodactylus* sp. We found that parasite infection increased the likelihood of a set of behaviors appearing, including maintaining higher nearest-neighbor distance, spending more time in isolation, transitioning to isolation from bigger group sizes, and occupying more peripheral and frontal positions across different group sizes. Overall, *Gyrodactylus* sp. infection does not seem to induce a single novel disease-specific behavior but rather increases the likelihood for a diseased individual to occupy a space in the extremes of the range of guppy behavioral repertoires.

Patterns of animal behavior change across timescales, from near-instantaneous fusion in response to increased risk of predation to hourly fission-fusion patterns, to longer timescale seasonal migration. Pathogen infection is an example of a process that has implications over a range of timescales, from short-term modification to locomotion patterns to isolation and avoidance, that might happen over minutes, hours, or days after initial exposure. Therefore, in Chapter 3 of this thesis, we investigated how, within a 2-hour interaction window, differential social influence within guppies shoals varies from the near-instantaneous time scale (1 second) to a relatively longer timescale (150 seconds). From this analysis, we inferred a form of leader-follower dynamics, evaluated how patterns of coordination among shoal members varied across populations, and how parasitism affected the role of the infected individual within the group. At a population level, we found that fish shoals use different rules to interact with each other depending on the timescale of the interaction, suggesting that different individuals are more, or less, influential over different timescales (i.e. those who lead short-timescale schooling dynamics are not those who influence the overall activity level exhibited by individuals in

the population). Moreover, we found that the magnitude of leadership scores decreased for increasing parasite loads, and that this was likely driven by an overall decrease in the correlation of the speed and activity patterns of the infected fish, with the rest of the group. Overall, different timescales of interaction can have different effects on the behavior of the group and this analysis has implications for understanding how individuals within a group make decisions across timescales, and how this varies with the presence of a contagious pathogen.

Disease spread is a major concern for public health, and understanding the factors that contribute to the spread of disease is critical for developing interventions to prevent or control outbreaks. In recent years, there has been increasing interest in the role of animal behavior in the spread of disease and we currently have tools that can aid a detailed analysis of how groups of animals move and interact. The study of disease spread is a complex and interdisciplinary field, and there is still much to learn about the factors that contribute to the spread of disease. With this thesis, we revealed that a multi-scale approach is necessary to understand the factors that contribute to the spread of disease in a social species, and we hope to see exciting developments in this direction in the future.

Appendices

Appendix A

Appendix to Chapter 1

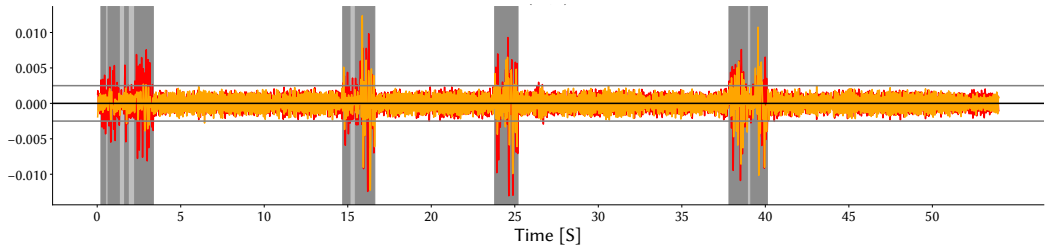


Figure A.1: **Identification of segments of pectoral fin use.** For each fish we estimated pectoral fin movement by calculating the normalized pectoral fin motion speed along the y-offset of each pectoral fin, separately (y-axis). To quantify the proportion of time pectoral fins were used, we defined epochs of consecutive frames within which pectoral fin velocity was above the arbitrary threshold of ± 0.0025 cm/s (solid gray horizontal lines). The minimum unsteady epoch duration length was set to 1/8 of a second. When two or more epochs were found to be separated by less than 0.5s, they were combined into one. Dark gray vertical bars denote windows where the data is above the threshold. The proportion of time the fish used pectoral fins was then calculated as the number of frames found with the filter divided by the total number of recorded frames for that flow speed. Light gray vertical bars show cases in which two windows of consecutive frames were separated by less than 0.5 seconds and thus merged into one larger segment.

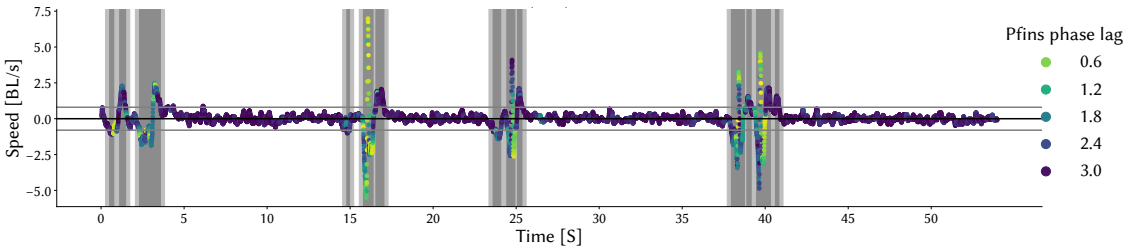


Figure A.2: **Identification of segments of unsteady swimming.** Time series of the fish speed at an example flow speed. To quantify the proportion of time spent in unsteady swimming, we defined epochs of consecutive frames within which the fish the fish's velocity was above the arbitrary threshold of ± 0.8 BL·s^{-1} (solid horizontal lines). The minimum unsteady epoch duration length was set to 1/8 of a second. When two or more epochs were found to be separated by less than 0.5s, they were combined into one. Each epoch was padded with 1/4s frames to account for the beginning and the end of an unsteady event. The proportion of time the fish spent in unsteady swimming was then calculated as the number of frames found with the filter divided by the total number of recorded frames for that flow speed. Dark gray vertical bars denote windows where the data is above the threshold. Light gray vertical bars show cases in which two windows of consecutive frames were separated by less than 0.5 seconds and thus merged into one larger segment. The figure also shows the variation of pectoral fins phase lag (color) at stable and unstable swimming states. This is purely demonstrative, where low values of phase lag correspond to symmetric use and higher values of phase lag correspond to alternate pectoral fin flapping.

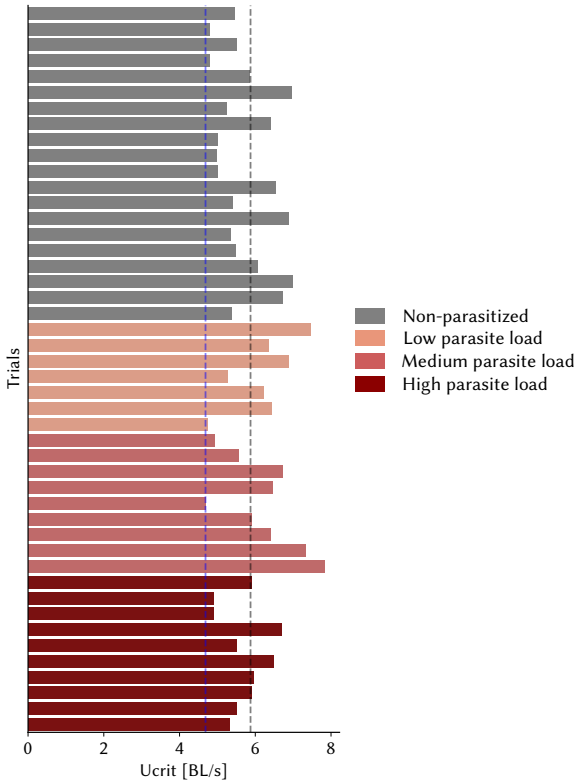


Figure A.3: U_{crit} **values for all fish tested.** Each horizontal bar shows the U_{crit} for each fish tested. The values are divided by parasite load group - including non-parasitized fish. The gray dotted vertical line marks the median U_{crit} of all fish tested. The blue vertical line marks the minimum U_{crit} recorded. Gray: Non-parasitized fish; Salmon: Low-parasite load; Coral: Medium-parasite load; Dark red: High-parasite load.

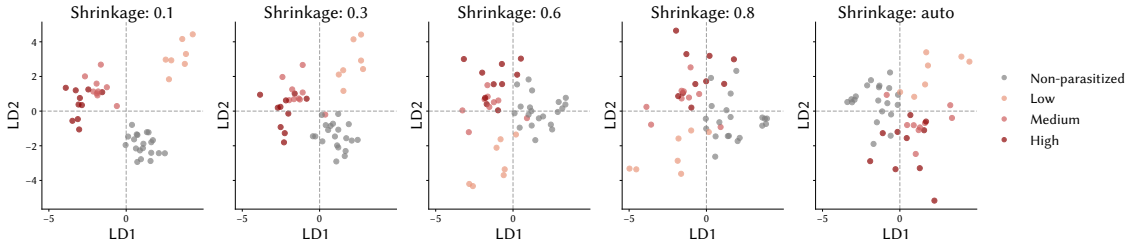
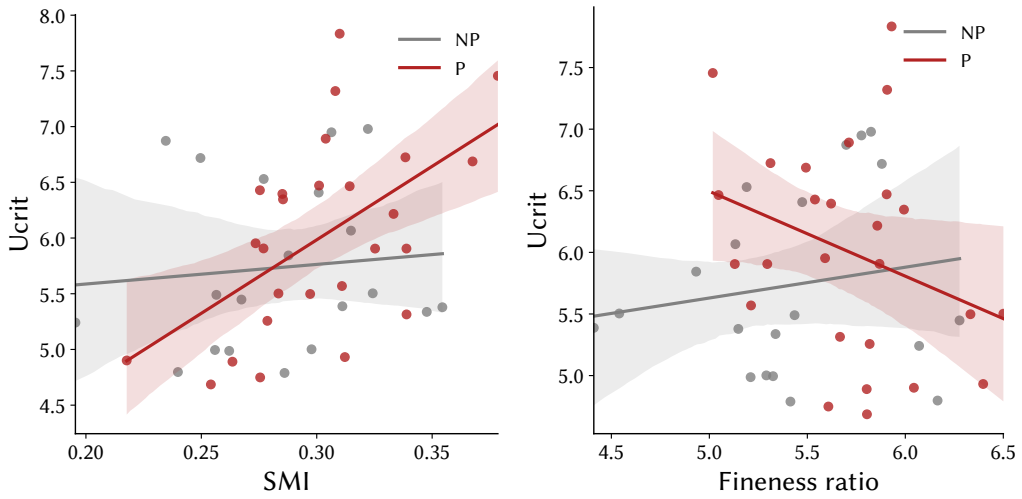


Figure A.4: **Effect of the shrinkage value on LDA.** Comparison of how the value of the shrinkage parameter can influence the LDA classification result. In our analysis, we used the 'auto' setting, which computes the shrinkage using the Ledoit-Wolf lemma (Ledoit and Wolf, 2004).



(a) SMI and U_{crit}

(b) Fineness and U_{crit}

Figure A.5: **Correlation between morphology and swimming performance.** Correlation between SMI (a) and Fineness ratio (b) with U_{crit}. (a) Non-parasitized fish show a weak correlation between SMI and U_{crit} (r=0.09), compared to parasitized fish, which show a strong positive correlation (r=0.54). (b) Non-parasitized fish show a weak correlation between Fineness ratio and U_{crit} (r=0.16), compared to parasitized fish, which show a moderate negative correlation (r=-0.31). NP: non-parasitized fish; P: parasitized fish. Line plots show the mean and 95% CI of the mean.

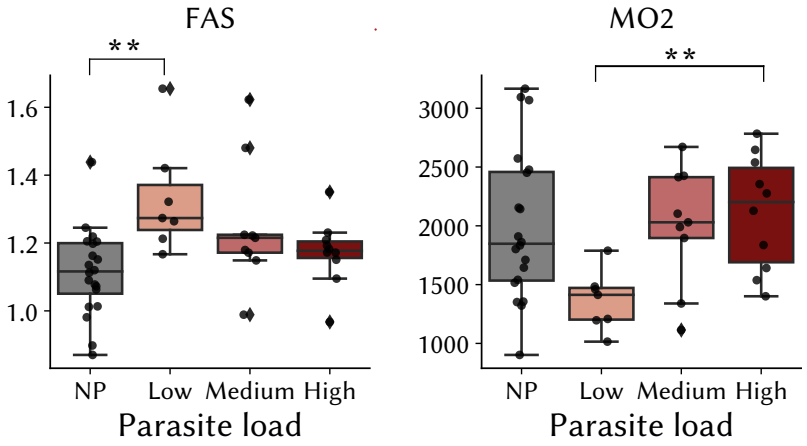


Figure A.6: **Average FAS and MO2 among treatments.** Values of FAS and MO2 as a function of parasite load (x-axis). FAS is significantly different among treatments (one-way ANOVA, p=0.006), with FAS values of non-parasitized fish significantly lower than low-parasite load individuals (post-hoc Tukey Honestly Significant Difference p=0.0014). Average MO2 (across flow speeds) is also significantly different among treatments (one-way ANOVA, p=0.0388), in particular between low-parasite load and high-parasite load fish (post-hoc Tukey Honestly Significant Difference p=0.0021, bonferroni corrected). Box plots give the median, interquartile range (box), and values that sit outside of 1.5 times the interquartile range (whiskers).

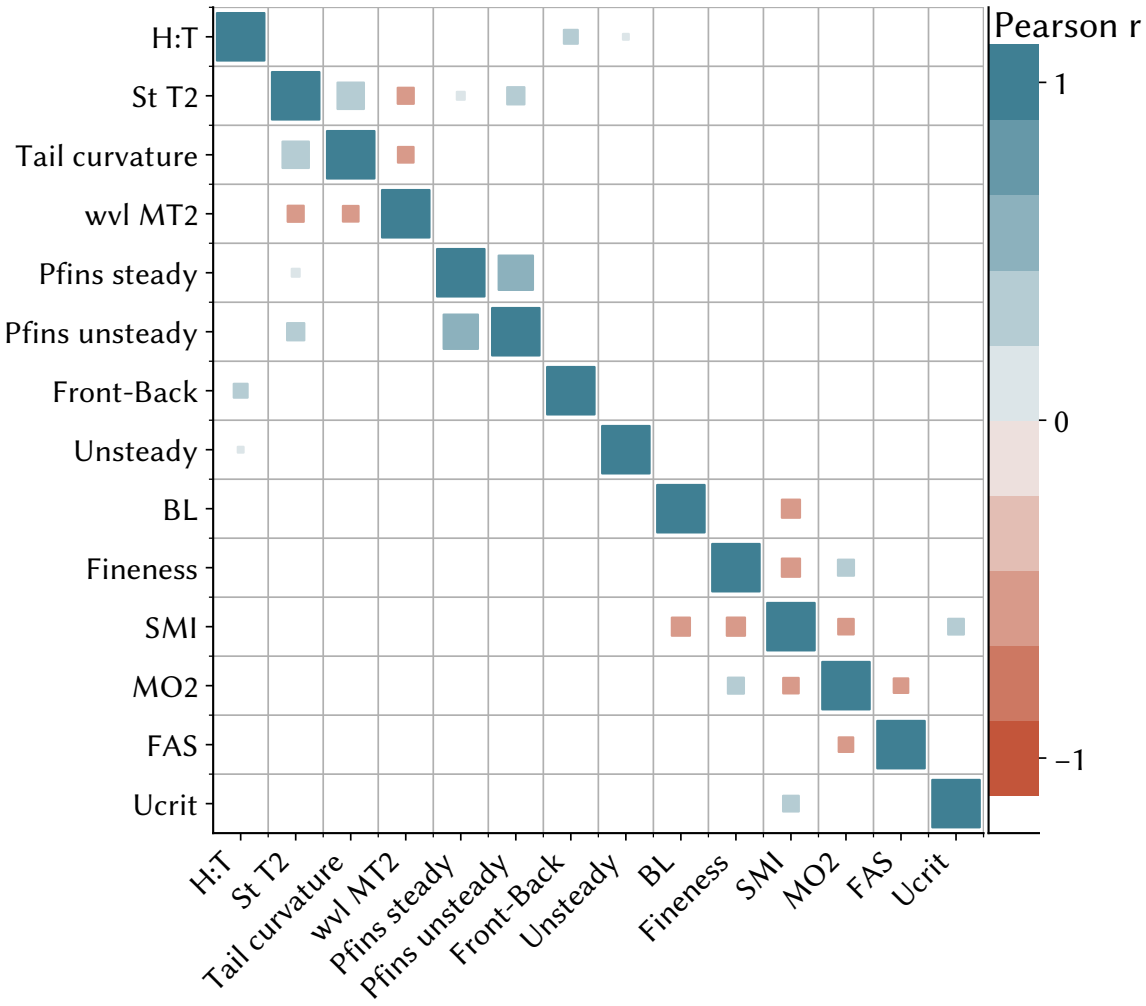


Figure A.7: **Correlation matrix between features.** The matrix shows how all features of kinematics, physiology, morphology, and U_{crit} correlate with each other. Positive correlations are shown in blue, negative correlations are shown in red.

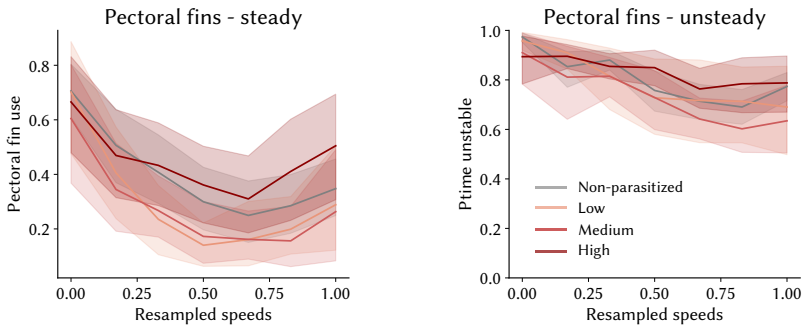


Figure A.8: **Pectoral fin use and parasite load.** Comparison of the pectoral fin use in steady (left) and unsteady (right) swimming states, comparing non-parasitised (gray), low, medium, and high parasite load fish. Line plots show the mean and 95% CI of the mean.

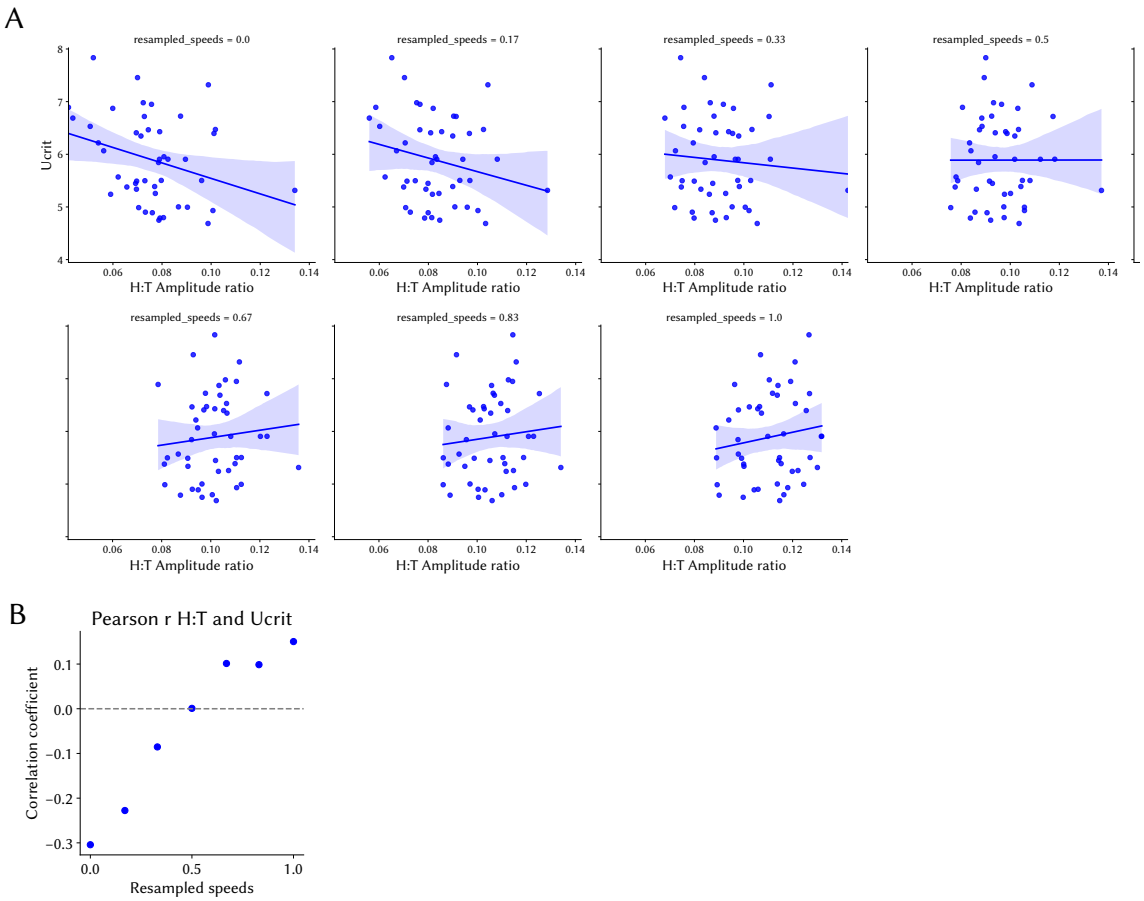


Figure A.9: **Correlation between H:T Amplitude ratio and U_{crit}** (A) Correlation between Head:Tail amplitude ratio and U_{crit} separated for each resampled flow speed (each single regression plot). (B) Pearson correlation values between H:T amplitude ratio and U_{crit} at each flow speed.

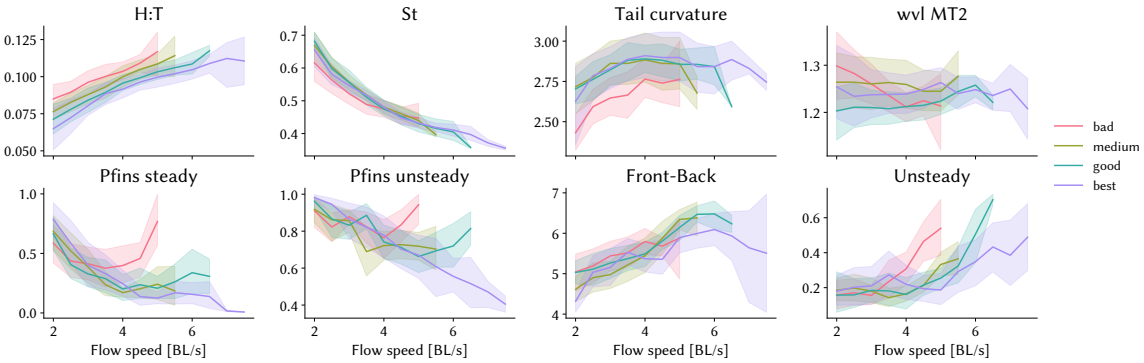


Figure A.10: **Distributions of kinematic measures as a function of performance group.** All kinematic features are shown as a function of real flow speed, after dividing the fish into four swimming performance categories. The categories are defined by the 25th, 50th, 75th quantile of all U_{crit} values of the tested population, categories are then renamed as 'bad', 'medium', 'good', and 'best' performers. For a list of the abbreviations in the titles, see Figure 1.4. Line plots show the mean and 95% CI of the mean.

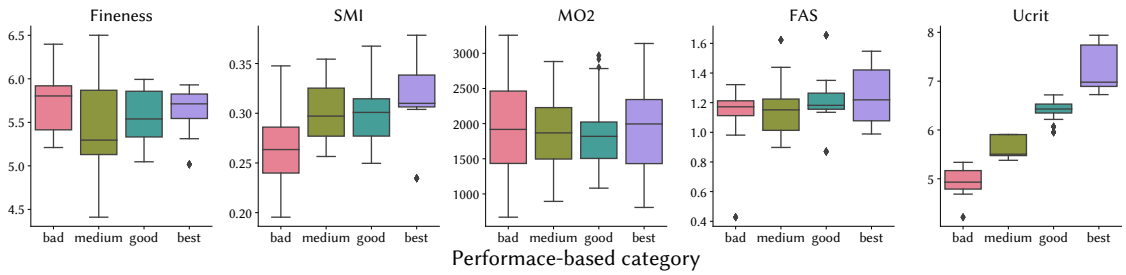


Figure A.11: **Distributions of morphology and physiology measures as a function of performance** Morphology (Fineness, SMI) and physiology (average MO2, FAS) are shown for each performance category. The categories are defined by the 25th, 50th, 75th quantile of all U_{crit} values of the tested population, categories are then renamed as 'bad', 'medium', 'good', and 'best' performers. Line plots show the mean and 95% CI of the mean. Box plots give the median, interquartile range (box), and values that sit outside of 1.5 times the interquartile range (whiskers).

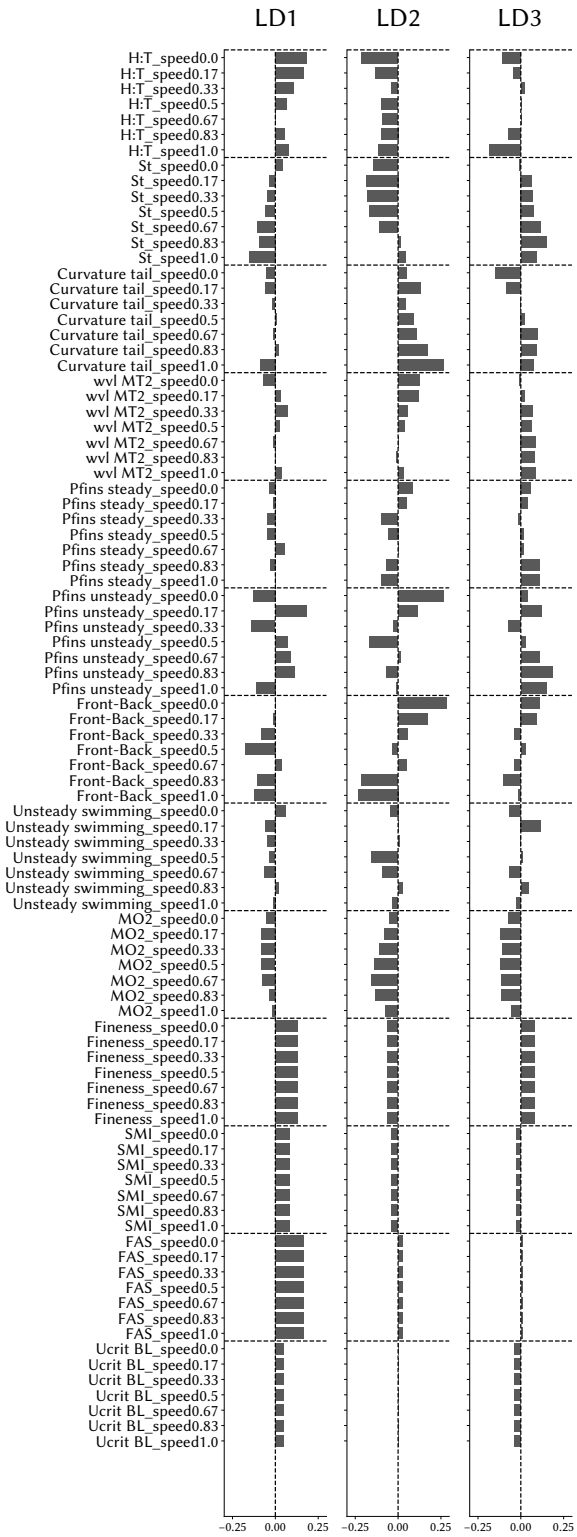


Figure A.12: **All the LDA components** All the weights of the first three LDA components. Each feature is included in the decomposition as a function of flow speed. On the y-axis, all features are listed in increasing order (top to bottom) as a function of interpolated flow speed. The naming convention of the features listed is as follows: *feature*_speed, followed by the speed value (between 0.0 and 1.0, see Section 1.2.8). Horizontal dotted lines separate the features.

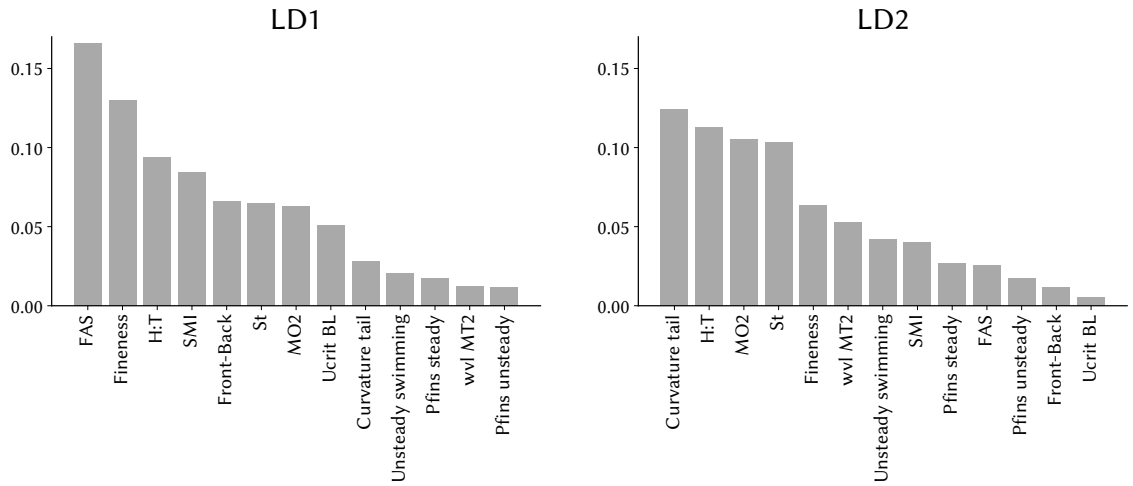


Figure A.13: **Features of LD1 and LD2 in decreasing order by weight.** Average weighting for each feature across flow speeds first two linear discriminants ordered by their absolute value to show the relative importance of each feature.

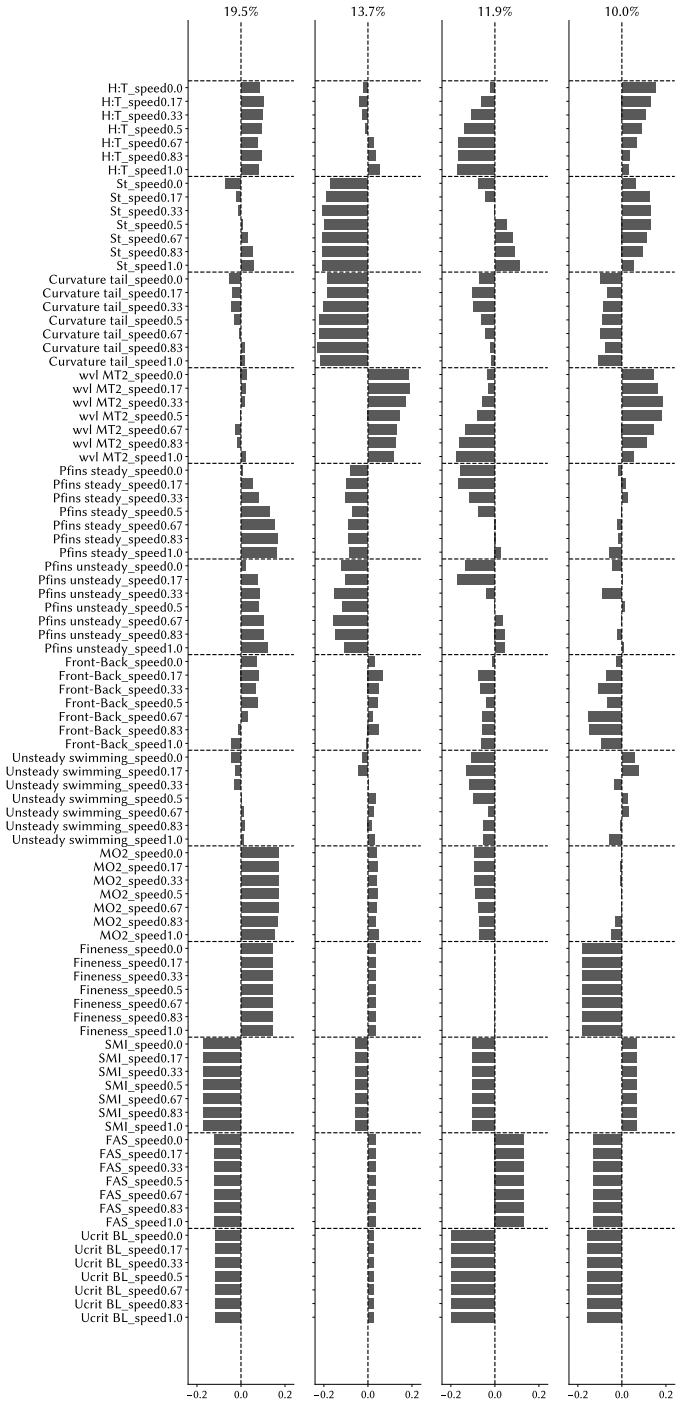
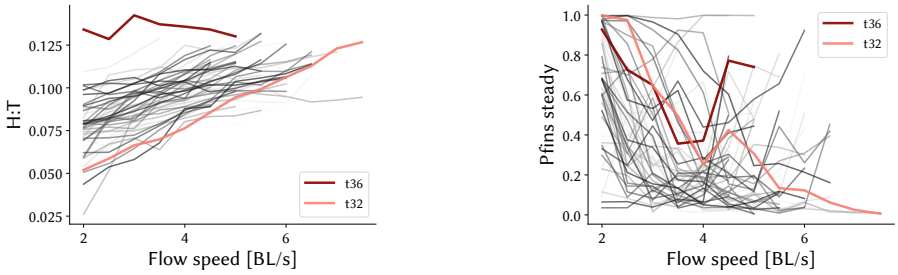


Figure A.14: **All the PCA components** All the weights of the first four PCA components. Positive/negative weightings in the PCA components represent higher/lower values of a metric with respect to trial average. Each feature is included in the decomposition as a function of flow speed (for a list of all the features see Figure 1.4). On the y-axis, all features are listed in increasing order (top to bottom) as a function of interpolated flow speed. The naming convention of the features listed is as follows: *feature*_speed, followed by the speed value (between 0.0 and 1.0, see Section 1.2.8). Horizontal dotted lines separate the features.



(a) H:T amplitude ratio for two example fish

(b) Pectoral fin use for two example fish

Figure A.15: **Illustrative example of kinematics differences between two parasitized fish.** To illustrate part of the variation we find among parasitized individuals in the tested population, we show the comparison of two paritized fish on two kinematics metrics: Head:tail (H:T) amplitude ratio (left) and Pectoral fin use at steady swimming (Pfins steady, right). The two compared fish are marked with two different colors, salmon and dark red. The data from the rest of the population (both non-parasitized and parasitized) is shown in gray. The largest H:T value measured in our data belongs to the individual with the largest number of parasites (approximately 250 worms) (left, t36 dark red color). The same fish also shows an increase in pectoral fin use in steady swimming at higher flow speeds (right, t36 dark red color). By contrast, we compared this individual to another parasitized fish (parasite load of 70 worms) that showed very low H:T compared to the other fish (right, t32 salmon color) at all flow speeds and constant decrease of pectoral fin use at high flow speeds (right, t32 salmon color). This fish also reached one of the highest U_{crit} and counted about 15 worms less compared to the load at the beginning of the experiment.

Appendix B

Appendix to Chapter 2

B.1 Control experiments with sham-infection

B.1.1 Sham-infection procedure

The experimental procedure described in section 2.2.3 includes manipulation of the fish, such as elastomer tagging, change of holding tank, anesthesia, semi-isolation. These procedures have the potential to stress a fish and possibly affect the behavior recorded during an experiment. In order to control for possible effects of the manipulation, we collected experimental trial recordings where we substituted the parasite infection procedure with a of sham-infection - thus without adding a parasite to the fish's skin. The rest of the experimental procedure is maintained identical to that described in 2.2.3. We recorded a total of 37 control trials, of which we fully processed 34 trials. A total of 3 trials were excluded from further processing due to death or poor body condition of the Focal-fish. At the end of the Baseline condition of the control experiment, the Focal-fish was anesthetized and underwent the so-called 'sham-infection', a procedure that mimicked the beginning of an infection, but without adding parasites to the fish's skin (see section 2.2.2). After the sham-infection, fish were kept in semi-isolation in the same holding conditions as for the original experiment.

B.1.2 Sham-infection results

We analyzed the data in the same way as for the original experiment (that included real parasite infection). In order to verify if the manipulation on the fish affected the behavior of the Focal-fish in the Treatment condition, we looked at the effect of sham-infection on two metrics: nearest-neighbor distance and the proportion of time fish spend in different group sizes when shoaling.

First, we find that sham-infection did not induce a change in the nearest-neighbor distance of the Focal-fish in the Treatment condition (Figure B.1). From this, we conclude that the increase in nearest-neighbor distance of a parasitized fish (Figure 2.11a) is not driven by side effects of the experimental manipulation procedure.

Next, we verified the effects of the infection manipulation on the normalized proportion of time spent in each group size by the Focal-fish. More specifically we wanted to verify that the increase in the isolation rates that we find on the Focal-fish in the Treatment condition following parasitic infection (Figure 2.13), are not driven by the experimental manipulation procedure. In this case, too, we find that the sham-infection did not have any effect on the distribution of the normalized proportion of time the Focal-fish spends in each group size (Figure B.2).

Overall, we conclude that the manipulation we used to experimentally infect fish, as

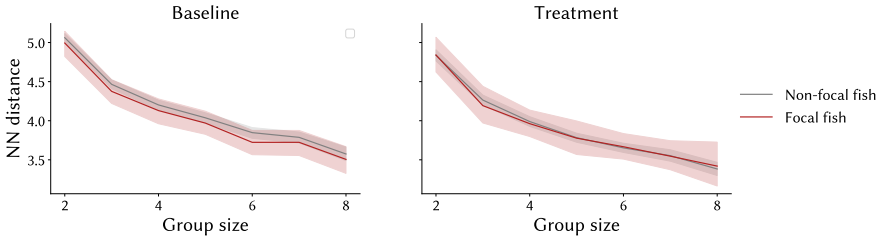


Figure B.1: **Control experiment: nearest-neighbor distance** The left figure shows data from the Baseline condition, compared to the Treatment condition shown on the right. Gray lines show pooled data of Non-focal fish, red lines show Focal-fish data. In both the Baseline and the Treatment condition we find no difference in the nearest-neighbor distance (cm) between the Non-focal and the Focal fish. This confirms that sham-infection did not alter nearest-neighbor distance of the Focal-fish after the manipulation procedure. Line plots show mean and 95% CI around the mean.

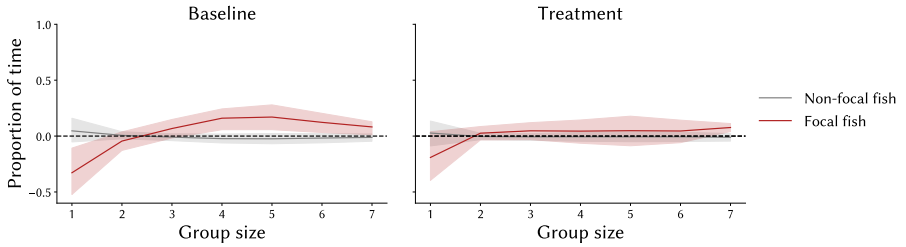


Figure B.2: **Control experiment: Normalized proportion of time spent in different group sizes.** Comparison between the normalized proportion of time Non-focal fish (gray) and Focal-fish (red) spend in the different group sizes at the Baseline (left) and at the Treatment (right) condition in the control experiment. Data from Non-focal fish is pooled together and the confidence interval of the mean is calculated across fish. In the Baseline condition, we see little to no change in the proportion of time NF and FF spend in different group sizes. The effect on decreased isolation of the Focal-fish is likely due to individual variation. In the Treatment condition, we find little to no difference between the Non-focal and the Focal-fish. This confirms that sham-infection did not alter the normalized proportion of time the Focal-fish spends in the different group sizes. Each group size is normalized and calculated as $(f - m)/m$ where f is an individual and m is the mean of all the fish in the trial. Line plots show mean and 95% CI around the mean.

well as the general experimental procedure that includes holding and experiment video recordings, did not induce any significant change in the behavior of the Focal-fish.

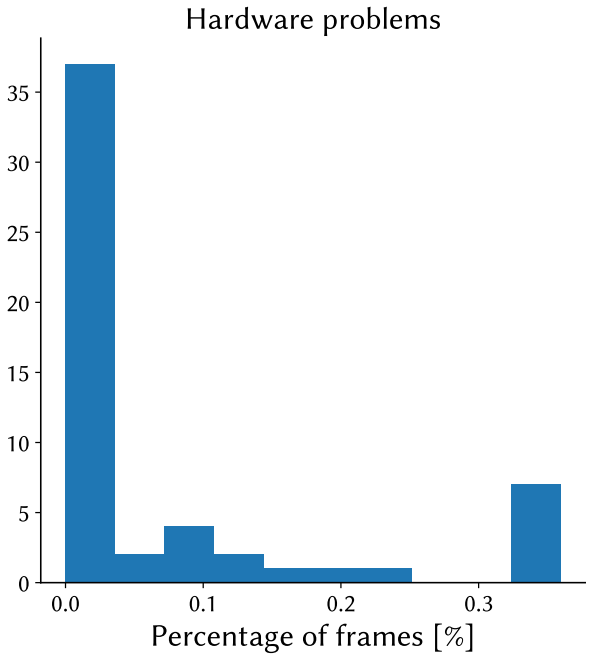


Figure B.3: **Discarded frames due to hardware problems.** Histogram of the percentage of the frames discarded due to hardware problems where the inter-frame interval was different from the 1/frame rate. The identified 'skipped' frames were assigned invalid numbers and excluded from further analysis.

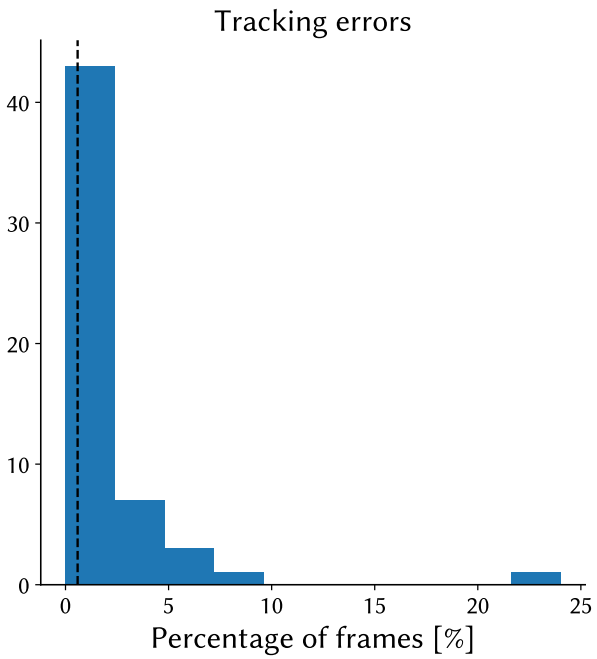
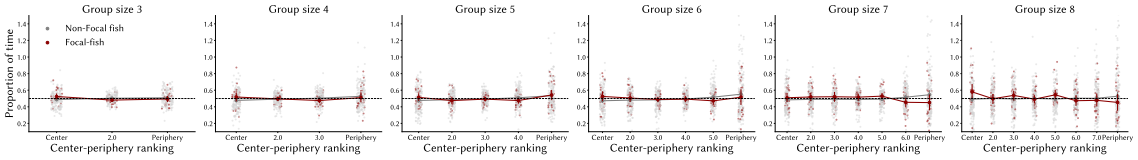
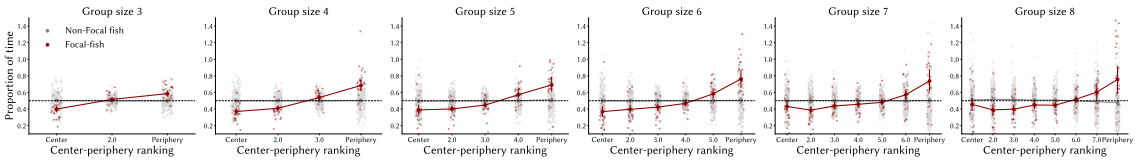


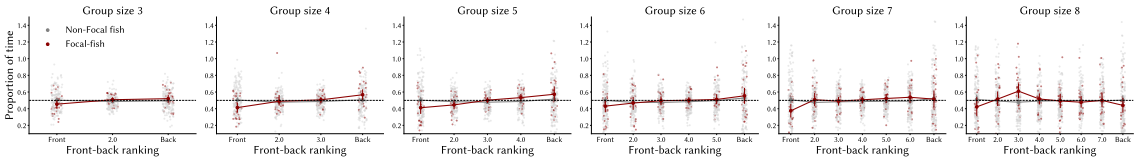
Figure B.4: **Discarded frames due to tracking errors.** Histogram of the percentage of the frames discarded due to tracking errors identified with the filter on speed described in section 2.2.4 and section 2.2.5. The dotted line shows the median of the distribution, that is 0.6%.



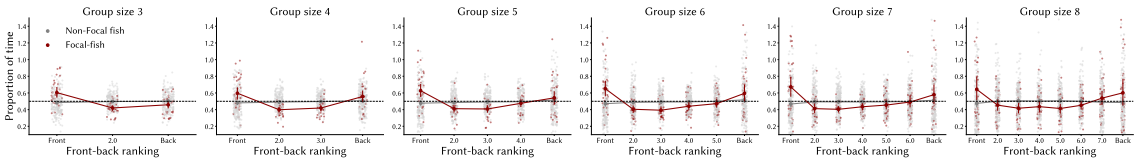
(a) Individual variation of time spent in Center-periphery ranking: Baseline condition



(b) Individual variation of time spent in Center-periphery ranking: Treatment condition



(c) Individual variation of time spent in Front-back ranking: Baseline condition



(d) Individual variation of time spent in Front-back ranking: Treatment condition

Figure B.5: **Proportion of time spent in each ranking position.** Each subplot (a-d) shows the normalized proportion of time each individual (one dot) spends in each ranking (x-axis) across group sizes from 3 to 8 (columns, left to right). (a,b) Normalized proportion of time each fish in the Baseline and Treatment condition, respectively, spend in each Center-periphery ranking across the different group sizes. (c,d) Normalized proportion of time each fish in the Baseline and Treatment condition, respectively, spent in each Front-back ranking across the different group sizes.

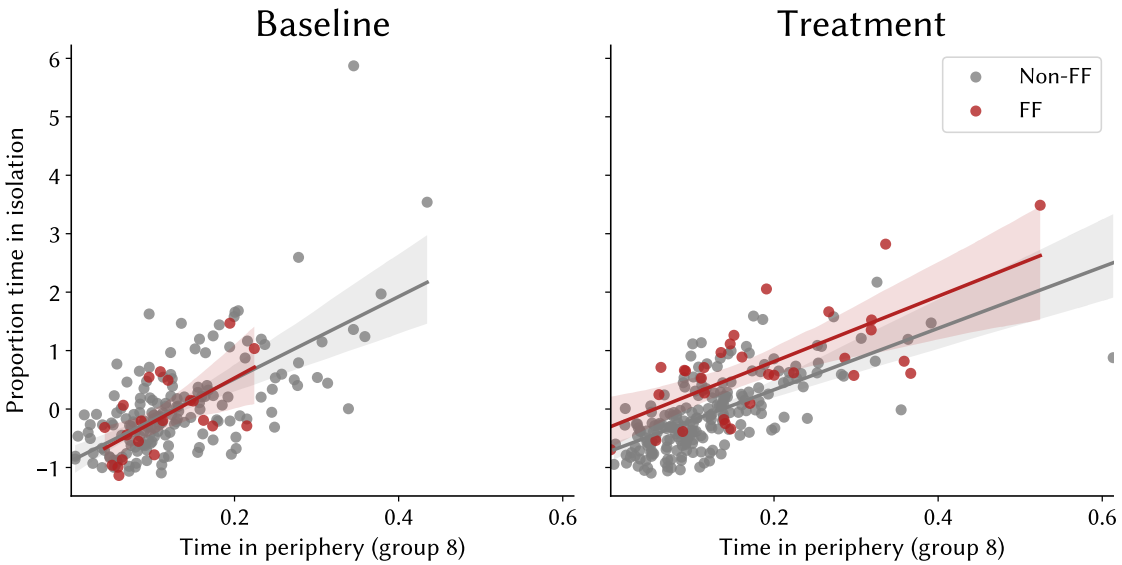


Figure B.6: **Correlation between isolation and peripheral position in the group.** Comparison between Baseline (left) and Treatment, for the Non-focal fish (gray) and the Focal-fish (red) of the correlation between the proportion of time spent in the most peripheral ranking position of a group size of 8 (x-axis), and the proportion of time spent in isolation (y-axis). Both NF and FF, in both Baseline and Treatment, show a positive correlation between the time spent in the periphery and the time spent in isolation. Non-FF: Non-Focal fish; FF: Focal-fish. All line plots give the mean and 95% CI of the mean at each group size.

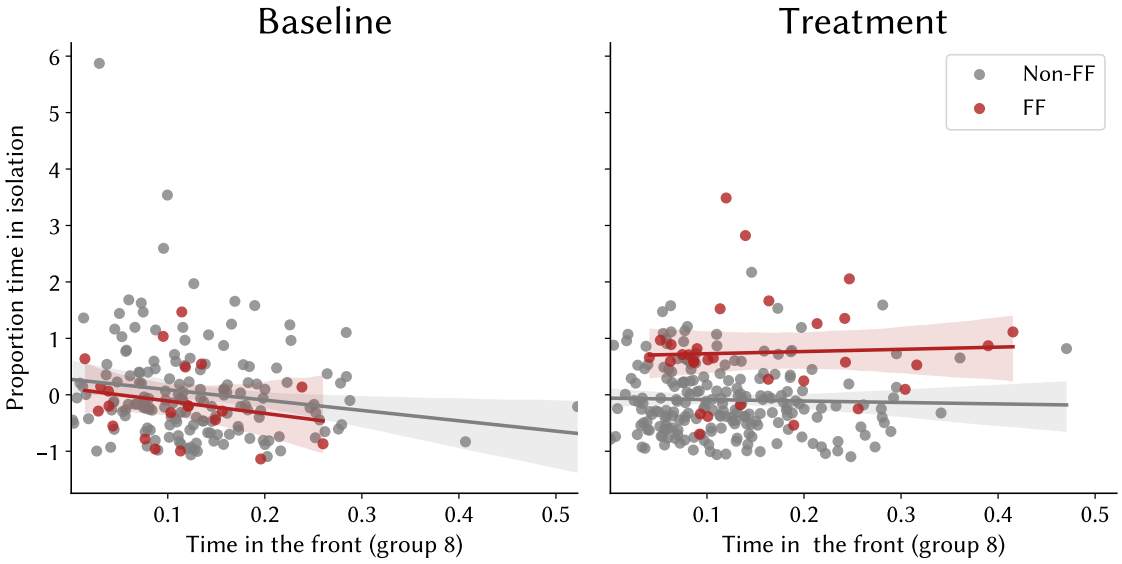


Figure B.7: **Correlation between isolation and frontal position in a group.** Comparison between Baseline (left) and Treatment, for the Non-focal fish (gray) and the Focal-fish (red) of the correlation between the proportion of time spent in the most frontal ranking position of a group size of 8 (x-axis), and the proportion of time spent in isolation (y-axis). We find no correlation between the time spent in the front of a group size of eight and the time spent in isolation. Non-FF: Non-Focal fish; FF: Focal-fish. All line plots give the mean and 95% CI of the mean at each group size.

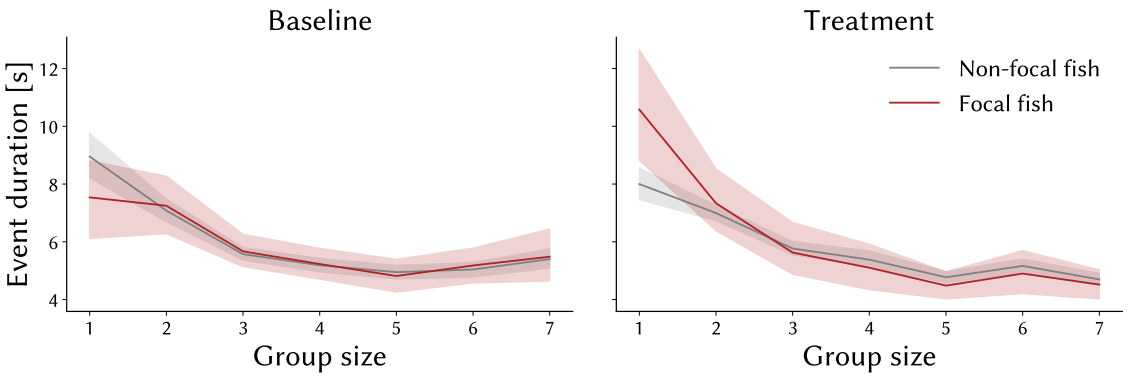


Figure B.8: **Mean event duration.** Comparison between Baseline (left) and Treatment (Right), for the Non-focal fish (gray) and the Focal-fish (red) of the average event duration across group sizes. Focal-fish in the Treatment condition show an average event duration of group size 1 that is higher than that of Non-focal fish in the same condition. The other group sizes do not show differences across conditions and fish. All line plots give the mean and 95% CI of the mean at each group size.

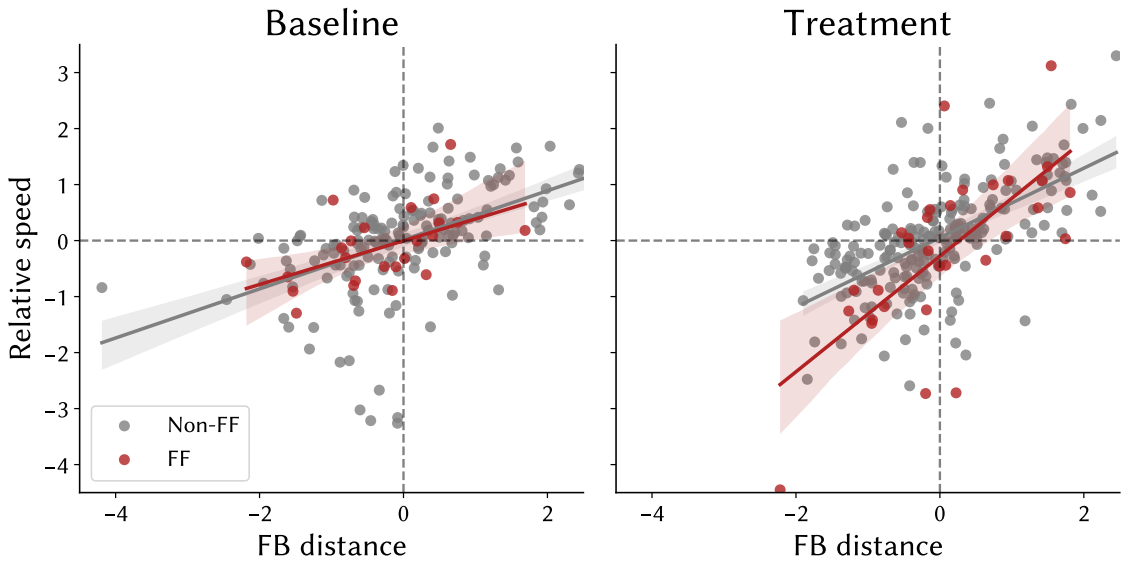


Figure B.9: **Speed relative to trial mean and relationship with front-back distance.** Correlation between front-back distance (x-axis) and relative speed (y-axis), for both the baseline condition (left) and the treatment condition (right). Front-back distance is positively correlated with the relative speed in all conditions and for both the non-focal fish (gray) and the focal-fish (red).

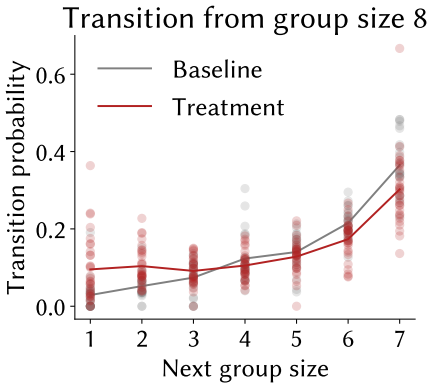
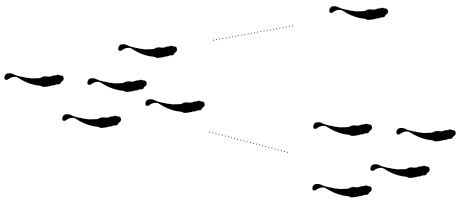
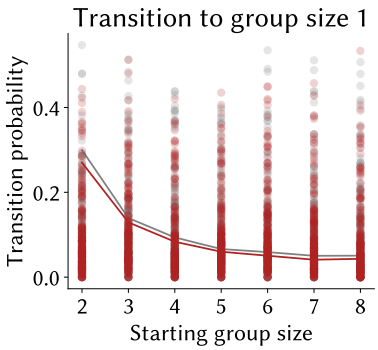


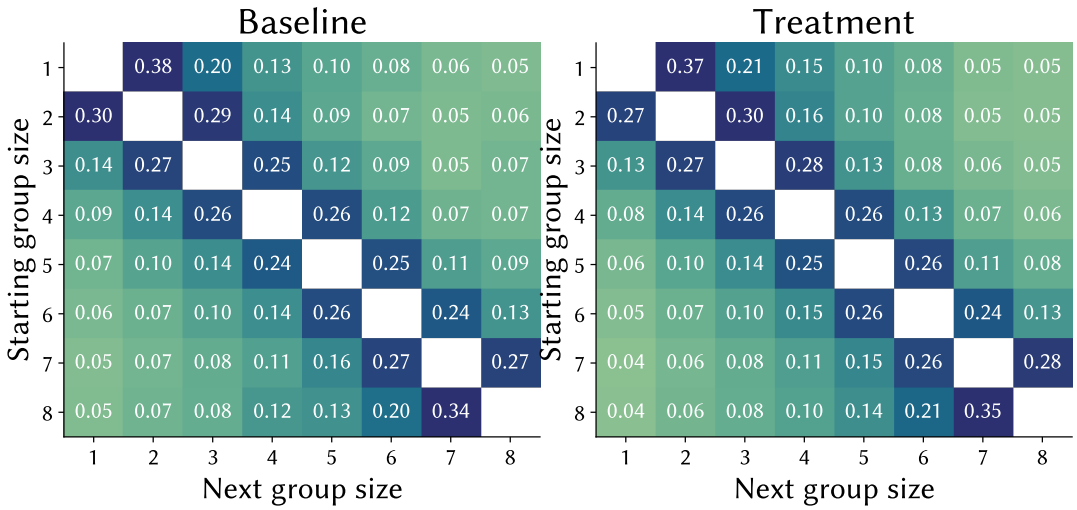
Figure B.10: **Transition probability from a group size of 8.** Transition probability of the focal-fish from the group size of 8 to any other group size (x-axis), in the baseline (gray) and the treatment (red) condition. Lines show the mean transition probability for the focal-fish in each group size comparing the two experimental conditions. The dots show data for each focal fish in the baseline (gray) and in the treatment (red) condition.



(a) Example transition from group size of 5 to groups of 4 and 1



(b) Transition probability 8 to 1



(c) Transition matrix for all combinations

Figure B.11: **Fission-fusion transition probability for the Non-focal fish.** (a) Example of fission from a group size of 5 to two smaller groups of 1 and 4 individuals. (b) Transition probability of the Non-focal fish from any group size (x-axis) to the group size of 1, in the baseline (gray) and the treatment (red) condition. Lines show the mean transition probability for the Focal-fish from any group size (2-8, x-axis) to the group size of 1. The dots show data for each Non-focal-fish in the Baseline (gray) and in the Treatment (red) condition. (c) Fission-fusion associated probability for the Non-focal fish of transitioning between any starting group size (y-axis), to any other group size (the next group to join, x-axis). The first column of each matrix refers to the probability of the Non-focal fish at the Baseline (left) and at the Treatment (right) to transition from any group size (2-8) to the group size of 1, as shown in 2.15b. Color denotes the transition probability, from low (light) to high probability (dark).

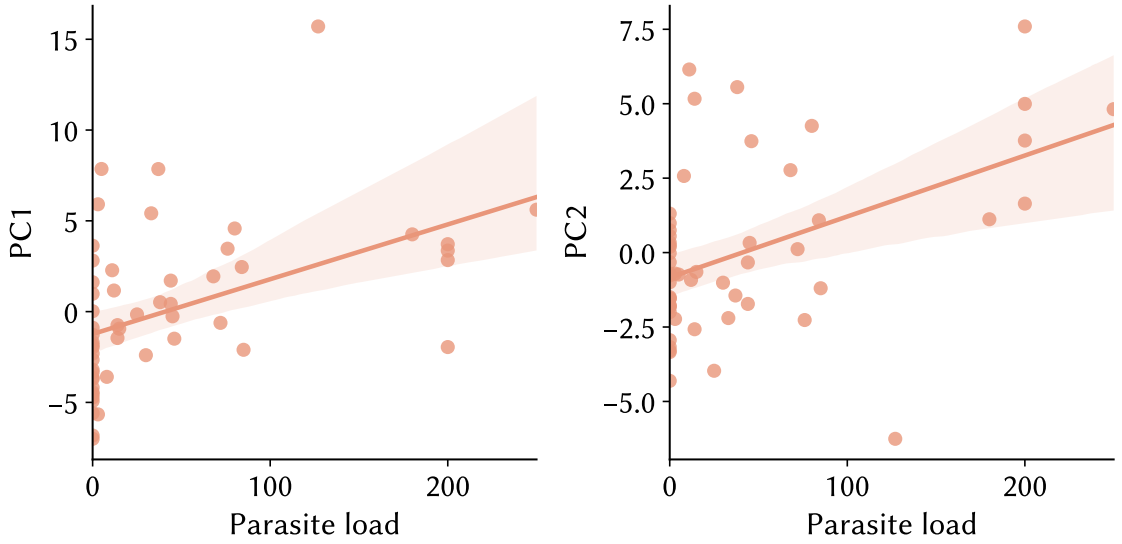


Figure B.12: **Correlation of Parasite load with PC1 and PC2.** Parasite load positively correlates with both PC1 and PC2 (r=0.47, p<0.001 both)

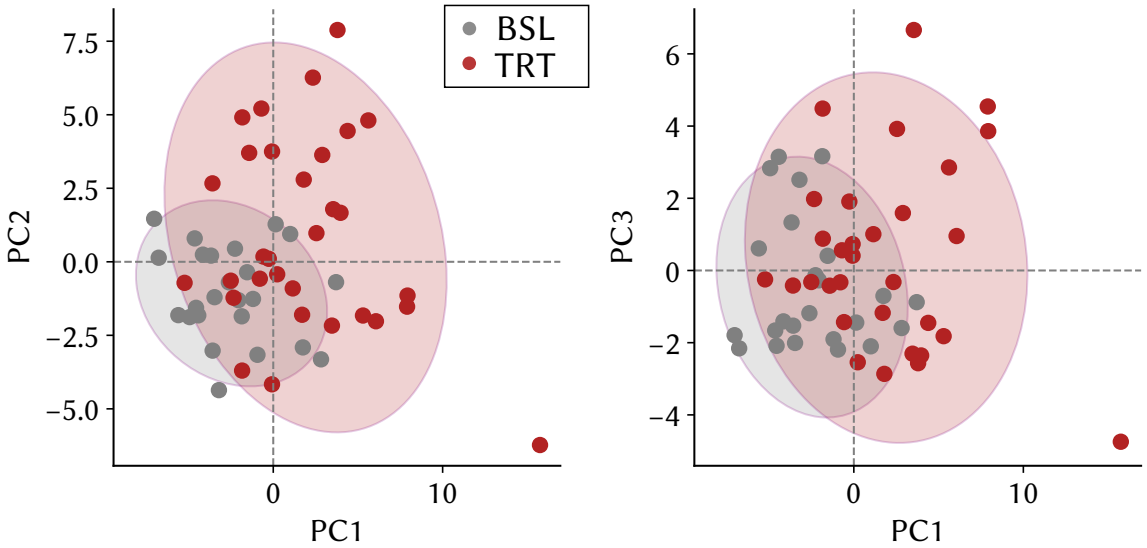


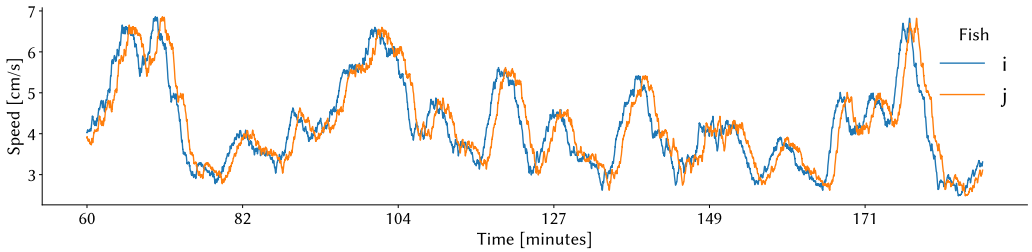
Figure B.13: **Projections onto the first 3 PCA modes, values for the Focal-fish, with confidence intervals.)** (a) Projections onto Principal components 1 and 2 for the Focal-fish in the Baseline (gray) and the Treatment (red) condition. (b) Projections onto Principal components 1 and 3 for the Focal-fish in the Baseline (gray) and the Treatment (red) condition. BSL: Baseline; TRT: Treatment. The ellipses show the 95% confidence interval around each principal component.

Appendix C

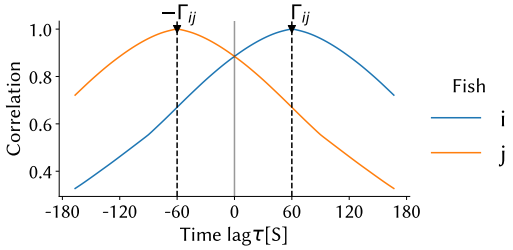
Appendix to Chapter 3

C.1 Correlation of time-shifted signals

In this section, we show the correct interpretation of the time lag Γ_{ij}^{w}, calculated on the time of maximum correlation M_{ij}^{w} for any pair of individuals i, j. Figure C.1a shows an example $s_i(w,t)$ and $s_j(w,t)$, where in this case $s_j(w,t)$ is a copy of $s_i(w,t)$ shifted in time by 60 seconds. We calculated the correlation value $C_{ij}(w,\tau)$ with the equation in 3.1. The value of Γ_{ij}^{w}, that is the τ_{ij} value at the maximum correlation M_{ij}^{w}, corresponds to the time shift, that is 60 seconds. Based on this convention, positive Γ_{ij}^{w} indicates that fish i is a 'leader' that initiates speed changes, whereas negative Γ_{ij}^{w} indicates that individual j is a 'follower', copying the speed changes with a time delay of τ of 60 seconds.



(a) Example of time-shifted speed traces



(b) Correlation values and Γ_{ij}^{w}

Figure C.1: **Method for calculating correlation of speed traces.** (a) Example s_i and s_j calculated with a moving average of 150s, where s_j is a copy of s_i shifted in time by 60 seconds (b) Correlation values C_{ij} between the traces s_i and s_j shown in (a) over a τ range of $\pm$ 1600 seconds. The value of Γ_{ij}, that is the τ_{ij} value at the maximum correlation M_{ij}, corresponds to the time shift, that is 60 seconds. Positive Γ_{ij} indicates that fish i is a 'leader' that initiates speed changes, whereas negative Γ_{ij} indicates that individual j is a 'follower', copying the speed changes with a time delay of τ of 60 seconds.

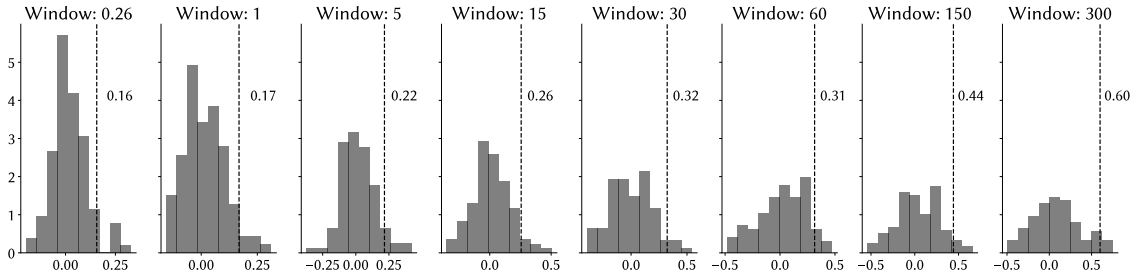


Figure C.2: **Maximum correlation across 100 random pairs for each window size.** To ensure the selection of pairs with correlation values that hold meaningful information about speed changes, we selected pairs with a $M_{ij}^{w} > M_{tr}^{w}$, depending on the window size. Each histogram shows the maximum correlation $C_{ij}(w, \tau)$ distribution for each window size w. The value of M_{tr}^{w} for chosen as the $95^{t}h$ percentile of each distribution.

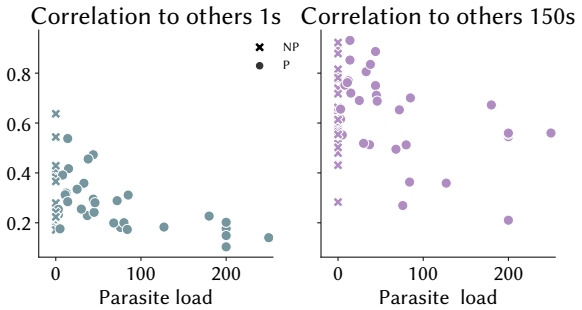


Figure C.3: **Correlation with others as a function of parasite load** (a) Correlation between parasite load and non-time-lagged pairwise correlation with the shoal members. Correlation values are generally smaller at the short timescale than they are at the long timescale (b). Overall, we find a general decrease in the correlation values with increasing parasite loads, both at the short (a) and the long (b) timescales. However, at the long timescale, we note that some of the high-parasite load individuals show a higher correlation to shoal members than it is found in the short timescale (a). Non-parasitized fish are marked with the symbol 'x', to distinguish from parasitized fish marked with the symbol 'o'.

References

Alexander, R. D. (1974). The Evolution of Social Behavior. *Annual Review of Ecology and Systematics*, 5(1):325–383. https://doi.org/10.1146/annurev.es.05.110174.001545.

Ashraf, I., Bradshaw, H., Ha, T.-T., Halloy, J., Godoy-Diana, R., and Thiria, B. (2017). Simple phalanx pattern leads to energy saving in cohesive fish schooling. *Proceedings of the National Academy of Sciences of the United States of America*, 114(36):9599–9604.

Bainbridge, R. (1958). The speed of swimming of fish as related to size and to the frequency and amplitude of the tail beat. *Journal of Experimental Biology*, page 25.

Bakke, T., Cable, J., and Harris, P. (2007). The Biology of Gyrodactylid Monogeneans: The "Russian-Doll Killers". In *Advances in Parasitology*, volume 64, pages 161–460. Elsevier.

Banet, A. I., Svendsen, J. C., Eng, K. J., and Reznick, D. N. (2016). Linking reproduction, locomotion, and habitat use in the Trinidadian guppy (Poecilia reticulata). *Oecologia*, 181(1):87–96.

Barber, I. and Dingemanse, N. J. (2010). Parasitism and the evolutionary ecology of animal personality. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365(1560):4077–4088.

Barber, I., Downey, L. C., and Braithwaite, V. A. (1998). Parasitism, oddity and the mechanism of shoal choice. *Journal of Fish Biology*, 53(6):1365–1368. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1095-8649.1998.tb00256.x.

Barber, I., Hoare, D., and Krause, J. (2000). Effects of parasites on fish behaviour: a review and evolutionary perspective. *Reviews in Fish Biology and Fisheries*, page 36.

Barber, I. and Huntingford, F. A. (1996). Parasite Infection Alters Schooling Behaviour: Deviant Positioning of Helminth-Infected Minnows in Conspecific Groups. *Proceedings: Biological Sciences*, 263(1374):1095–1102. Publisher: The Royal Society.

Barber, I., Mora, A. B., Payne, E. M., Weinersmith, K. L., and Sih, A. (2017). Parasitism, personality and cognition in fish. *Behavioural Processes*, 141:205–219.

Behringer, D. C., Butler, M. J., and Shields, J. D. (2006). Ecology: avoidance of disease by social lobsters. *Nature*, 441(7092):421.

Behringer, D. C., Karvonen, A., and Bojko, J. (2018). Parasite avoidance behaviours in aquatic environments. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 373(1751):20170202.

Best, A., White, A., and Boots, M. (2008). Maintenance of host variation in tolerance to pathogens and parasites. *Proceedings of the National Academy of Sciences*, 105(52):20786–20791.

Binning, S. A., Roche, D. G., and Fulton, C. J. (2014). Localised intraspecific variation in the swimming phenotype of a coral reef fish across different wave exposures. *Oecologia*, 174(3):623–630.

Binning, S. A., Roche, D. G., and Layton, C. (2013). Ectoparasites increase swimming costs in a coral reef fish. *Biology Letters*, 9(1):20120927.

Bordes, F., Blumstein, D. T., and Morand, S. (2007). Rodent sociality and parasite diversity. *Biology Letters*, 3(6):692–694. Publisher: Royal Society.

Brett, J. R. (1964). The Respiratory Metabolism and Swimming Performance of Young Sockeye Salmon. *Journal of the Fisheries Research Board of Canada*, 21(5):1183–1226.

Buck, J. C., Weinstein, S. B., and Young, H. S. (2018). Ecological and Evolutionary Consequences of Parasite Avoidance. *Trends in Ecology & Evolution*, 33(8):619–632. Publisher: Elsevier.

Buhl, J., Sumpter, D. J. T., Couzin, I. D., Hale, J. J., Despland, E., Miller, E. R., and Simpson, S. J. (2006). From Disorder to Order in Marching Locusts. *Science*, 312(5778):1402–1406.

Cable, J., Scott, E. C. G., Tinsley, R. C., and Harris, P. D. (2002). Behavior Favoring Transmission in the Viviparous Monogenean Gyrodactylus turnbulli. *Journal of Parasitology*, 88(1):183–184. Publisher: American Society of Parasitologists.

Cable, J. and van Oosterhout, C. (2007). The impact of parasites on the life history evolution of guppies (Poecilia reticulata): The effects of host size on parasite virulence. *International Journal for Parasitology*, 37(13):1449–1458.

Clark, T. D., Sandblom, E., and Jutfelt, F. (2013). Aerobic scope measurements of fishes in an era of climate change: respirometry, relevance and recommendations. *Journal of Experimental Biology*, 216(15):2771–2782.

Combes, C. (2001). *Parasitism: The Ecology and Evolution of Intimate Interactions*. University of Chicago Press. Google-Books-ID: LovrfCYloxgC.

Conradt, L., Krause, J., Couzin, I. D., and Roper, T. J. (2009). "Leading According to Need" in Self-Organizing Groups. *The American Naturalist*, 173(3):304–312. Publisher: The University of Chicago Press.

Conradt, L. and Roper, T. J. (2003). Group decision-making in animals. *Nature*, 421(6919):155–158. Number: 6919 Publisher: Nature Publishing Group.

Couzin, I. (2007). Collective minds. *Nature*, 445(7129):715–715.

Couzin, I. and Krause, J. (2003). Self-Organization and Collective Behavior in Vertebrates. *Advances in The Study of Behavior*, 32:1–75.

Couzin, I. D., Ioannou, C. C., Demirel, G., Gross, T., Torney, C. J., Hartnett, A., Conradt, L., Levin, S. A., and Leonard, N. E. (2011). Uninformed Individuals Promote Democratic Consensus in Animal Groups. *Science*, 334(6062):1578–1580.

Couzin, I. D., Krause, J., Franks, N. R., and Levin, S. A. (2005). Effective leadership and decision-making in animal groups on the move. *Nature*, 433(7025):513–516. Number: 7025 Publisher: Nature Publishing Group.

Couzin, I. D., Krause, J., James, R., Ruxton, G. D., and Franks, N. R. (2002). Collective Memory and Spatial Sorting in Animal Groups. *Journal of Theoretical Biology*, 218(1):1–11.

Couzin, I. D. and Laidre, M. E. (2009). Fission–fusion populations. *Current Biology*, 19(15):R633–R635.

Cremer, S., Armitage, S. A. O., and Schmid-Hempel, P. (2007). Social immunity. *Current biology: CB*, 17(16):R693–702.

Croft, D. P., Arrowsmith, B. J., Bielby, J., Skinner, K., White, E., Couzin, I. D., Magurran, A. E., Ramnarine, I., and Krause, J. (2003). Mechanisms underlying shoal composition in the Trinidadian guppy, Poecilia reticulata. *Oikos*, 100(3):429–438. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1034/j.1600-0706.2003.12023.x.

Croft, D. P., Edenbrow, M., Darden, S. K., Ramnarine, I. W., van Oosterhout, C., and Cable, J. (2011). Effect of gyrodactylid ectoparasites on host behaviour and social network structure in guppies Poecilia reticulata. *Behavioral Ecology and Sociobiology*, 65(12):2219–2227.

Curtis, V. A. (2014). Infection-avoidance behaviour in humans and other animals. *Trends in Immunology*, 35(10):457–464.

Di Santo, V., Goerig, E., Wainwright, D. K., Akanyeti, O., Liao, J. C., Castro-Santos, T., and Lauder, G. V. (2021). Convergence of undulatory swimming kinematics across a diversity of fishes. *Proceedings of the National Academy of Sciences of the United States of America*, 118(49):e2113206118.

Di Santo, V., Kenaley, C. P., and Lauder, G. V. (2017). High postural costs and anaerobic metabolism during swimming support the hypothesis of a U-shaped metabolism–speed curve in fishes. *Proceedings of the National Academy of Sciences*, 114(49):13048–13053. Publisher: Proceedings of the National Academy of Sciences.

Dobson, A., Lafferty, K. D., Kuris, A. M., Hechinger, R. F., and Jetz, W. (2008). Homage to Linnaeus: How many parasites? How many hosts? *Proceedings of the National Academy of Sciences*, 105(supplement_1):11482–11489. Publisher: Proceedings of the National Academy of Sciences.

Domenici, P., Turesson, H., Brodersen, J., and Brönmark, C. (2008). Predator-induced morphology enhances escape locomotion in crucian carp. *Proceedings of the Royal Society B: Biological Sciences*, 275(1631):195–201. Publisher: Royal Society.

Dugatkin, L. A., FitzGerald, G. J., and Lavoie, J. (1994). Juvenile three-spined sticklebacks avoid parasitized conspecifics. *Environmental Biology of Fishes*, 39(2):215–218.

Dyer, J. R., Johansson, A., Helbing, D., Couzin, I. D., and Krause, J. (2009). Leadership, consensus decision making and collective behaviour in humans. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1518):781–789. Publisher: Royal Society.

Griffiths, S. W. and Magurran, A. E. (1998). Sex and schooling behaviour in the Trinidadian guppy. *Animal Behaviour*, 56(3):689–693.

Hale, M. E., Day, R. D., Thorsen, D. H., and Westneat, M. W. (2006). Pectoral fin coordination and gait transitions in steadily swimming juvenile reef fishes. *Journal of Experimental Biology*, 209(19):3708–3718.

Halsey, L. G., Killen, S. S., Clark, T. D., and Norin, T. (2018). Exploring key issues of aerobic scope interpretation in ectotherms: absolute versus factorial. *Reviews in Fish Biology and Fisheries*, 28(2):405–415.

Hammer, C. (1995). Fatigue and exercise tests with fish. *Comparative Biochemistry and Physiology Part A: Physiology*, 112(1):1–20.

Harris, P. D. and Lyles, A. M. (1992). Infections of Gyrodactylus bullatarudis and Gyrodactylus turnbulli on Guppies (Poecilia reticulata) in Trinidad. *The Journal of Parasitology*, 78(5):912–914. Publisher: [The American Society of Parasitologists, Allen Press].

Hawley, D. M., Gibson, A. K., Townsend, A. K., Craft, M. E., and Stephenson, J. F. (2021). Bidirectional interactions between host social behaviour and parasites arise through ecological and evolutionary processes. *Parasitology*, 148(3):274–288. Publisher: Cambridge University Press.

Herbert-Read, J. E., Krause, S., Morrell, L. J., Schaerf, T. M., Krause, J., and Ward, A. J. W. (2013). The role of individuality in collective group movement. *Proceedings of the Royal Society B: Biological Sciences*, 280(1752):20122564.

Herbert-Read, J. E., Wade, A. S. I., Ramnarine, I. W., and Ioannou, C. C. (2019). Collective decision-making appears more egalitarian in populations where group fission costs are higher. *Biology Letters*, 15(12):20190556.

Hoare, D., Couzin, I., Godin, J.-G., and Krause, J. (2004). Context-dependent group size choice in fish. *Animal Behaviour*, 67(1):155–164.

Hoare, D., Reeves, P., and Krause, J. (1998). Positioning Behaviour in Roach Shoals: The Role of Body Length and Nutritional State. *Behaviour*, 135(8):1031–1039. Publisher: Brill.

Hockley, F. A., Wilson, C. A. M. E., Brew, A., and Cable, J. (2014). Fish responses to flow velocity and turbulence in relation to size, sex and parasite load. *Journal of The Royal Society Interface*, 11(91):20130814.

Houde, A. E. and Torio, A. J. (1992). Effect of parasitic infection on male color pattern and female choice in guppies. *Behavioral Ecology*, 3(4):346–351.

Ioannou, C. C., Singh, M., and Couzin, I. D. (2015). Potential Leaders Trade Off Goal-Oriented and Socially Oriented Behavior in Mobile Animal Groups. *The American Naturalist*, 186(2):284–293. Publisher: The University of Chicago Press.

Jolles, J. W., Boogert, N. J., Sridhar, V. H., Couzin, I. D., and Manica, A. (2017). Consistent Individual Differences Drive Collective Behavior and Group Functioning of Schooling Fish. *Current Biology*, 27(18):2862–2868.e7.

Jolles, J. W., King, A. J., and Killen, S. S. (2020a). The Role of Individual Heterogeneity in Collective Animal Behaviour. *Trends in Ecology & Evolution*, 35(3):278–291.

Jolles, J. W., Mazué, G. P. F., Davidson, J., Behrmann-Godel, J., and Couzin, I. D. (2020b). Schistocephalus parasite infection alters sticklebacks' movement ability and thereby shapes social interactions. *Scientific Reports*, 10(1):12282.

Katz, Y., Tunstrøm, K., Ioannou, C. C., Huepe, C., and Couzin, I. D. (2011). Inferring the structure and dynamics of interactions in schooling fish. *Proceedings of the National Academy of Sciences*, 108(46):18720–18725.

Kavaliers, M. and Choleris, E. (2018). The role of social cognition in parasite and pathogen avoidance. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 373(1751):20170206.

Kavaliers, M., Choleris, E., Ågmo, A., and Pfaff, D. W. (2004). Olfactory-mediated parasite recognition and avoidance: linking genes to behavior. *Hormones and Behavior*, 46(3):272–283.

Kennedy, C. E. J., Endler, J. A., Poynton, S. L., and McMinn, H. (1987). Parasite load predicts mate choice in guppies. *Behavioral Ecology and Sociobiology*, 21(5):291–295.

Kern, E. M. A. and Langerhans, R. B. (2019). Urbanization Alters Swimming Performance of a Stream Fish. *Frontiers in Ecology and Evolution*, 6.

Kern, P., Cramp, R. L., Gordos, M. A., Watson, J. R., and Franklin, C. E. (2018). Measuring U_{crit} and endurance: equipment choice influences estimates of fish swimming performance. *Journal of Fish Biology*, 92(1):237–247.

Kiesecker, J. M., Skelly, D. K., Beard, K. H., and Preisser, E. (1999). Behavioral reduction of infection risk. *Proceedings of the National Academy of Sciences*, 96(16):9165–9168. Publisher: Proceedings of the National Academy of Sciences.

Killen, S. S., Croft, D. P., Salin, K., and Darden, S. K. (2016). Male sexually coercive behaviour drives increased swimming efficiency in female guppies. *Functional Ecology*, 30(4):576–583. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/1365-2435.12527.

King, A. J. and Cowlishaw, G. (2007). When to use social information: the advantage of large group size in individual decision making. *Biology Letters*, 3(2):137–139. Publisher: Royal Society.

Korsmeyer, K. E., Steffensen, J. F., and Herskin, J. (2002). Energetics of median and paired fin swimming, body and caudal fin swimming, and gait transition in parrotfish (*Scarus schlegeli*) and triggerfish (*Rhinecanthus aculeatus*). *Journal of Experimental Biology*, 205(9):1253–1263.

Krause, J. (1993). The influence of hunger on shoal size choice by three-spined sticklebacks, Gasterosteus aculeatus. *Journal of Fish Biology*, 43(5):775–780. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1095-8649.1993.tb01154.x.

Krause, J., Bumann, D., and Todt, D. (1992). Relationship between the position preference and nutritional state of individuals in schools of juvenile roach (Rutilus rutilus). *Behavioral Ecology and Sociobiology*, 30(3-4):177–180.

Krause, J. and Godin, J.-G. J. (1996). Influence of Parasitism on Shoal Choice in the Banded Killifish (Fundulus diaphanus, Teleostei, Cyprinodontidae). *Ethology*, 102(1):40–49. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1439-0310.1996.tb01102.x.

Krause, J., Godin, J.-G. J., and Brown, D. (1996). Phenotypic Variability within and between Fish Shoals. *Ecology*, 77(5):1586–1591. Publisher: Ecological Society of America.

Krause, J., Hoare, D., Krause, S., Hemelrijk, C. K., and Rubenstein, D. I. (2000). Leadership in fish shoals. *Fish and Fisheries*, 1(1):82–89.

Lafferty, K. D. and Morris, A. K. (1996). Altered Behavior of Parasitized Killifish Increases Susceptibility to Predation by Bird Final Hosts. *Ecology*, 77(5):1390–1397. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.2307/2265536.

Langerhans, R. B. (2008). Predictability of phenotypic differentiation across flow regimes in fishes. *Integrative and Comparative Biology*, 48(6):750–768.

Langerhans, R. B. (2009). Trade-off between steady and unsteady swimming underlies predator-driven divergence in Gambusia affinis. *Journal of Evolutionary Biology*, 22(5):1057–1075. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1420-9101.2009.01716.x.

Lauder, G. V. (2015). Fish Locomotion: Recent Advances and New Directions. *Annual Review of Marine Science*, 7(1):521–545. _eprint: https://doi.org/10.1146/annurev-marine-010814-015614.

Lauer, J., Zhou, M., Ye, S., Menegas, W., Schneider, S., Nath, T., Rahman, M. M., Di Santo, V., Soberanes, D., Feng, G., Murthy, V. N., Lauder, G., Dulac, C., Mathis, M. W., and Mathis, A. (2022). Multi-animal pose estimation, identification and tracking with DeepLabCut. *Nature Methods*, 19(4):496–504.

Ledoit, O. and Wolf, M. (2004). A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis*, 88(2):365–411.

Leeper, T. J. (2017). Interpreting Regression Results using Average Marginal Effects with R's margins. page 31.

Li, L., Nagy, M., Graving, J. M., Bak-Coleman, J., Xie, G., and Couzin, I. D. (2020). Vortex phase matching as a strategy for schooling in robots and in fish. *Nature Communications*, 11(1):5408.

Liao, J. C. (2007). A review of fish swimming mechanics and behaviour in altered flows. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 362(1487):1973–1993.

Lopes, P. C. (2022). Anticipating infection: How parasitism risk changes animal physiology. *Functional Ecology*.

Lopes, P. C., Block, P., and König, B. (2016). Infection-induced behavioural changes reduce connectivity and the potential for disease spread in wild mice contact networks. *Scientific Reports*, 6(1):31790. Number: 1 Publisher: Nature Publishing Group.

Lopes, P. C., French, S. S., Woodhams, D. C., and Binning, S. A. (2021). Sickness behaviors across vertebrate taxa: proximate and ultimate mechanisms. *Journal of Experimental Biology*, 224(9):jeb225847.

Lopes, P. C., French, S. S., Woodhams, D. C., and Binning, S. A. (2022). Infection avoidance behaviors across vertebrate taxa: Patterns, processes, and future directions. In *Animal Behavior and Parasitism*, page 22. Oxford University Press.

Maceda-Veiga, A., Green, A., and Sostoa, A. (2014). Scaled mass index shows how habitat quality influences the condition of four fish taxa in north-eastern Spain, and provides a novel indicator of ecosystem health. *Freshwater Biology*, 59.

Magurran, A. E. (2005). *Evolutionary Ecology: The Trinidadian Guppy*. OUP Oxford. Google-Books-ID: 2WxwxwSlhXYC.

Magurran, A. E., Pitcher, T. J., and Dodd, J. M. (1987). Provenance, shoal size and the sociobiology of predator-evasion behaviour in minnow shoals. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, 229(1257):439–465. Publisher: Royal Society.

Marcogliese, D. J. (2004). Parasites: Small Players with Crucial Roles in the Ecological Theater. *EcoHealth*, 2(1):151–164.

Marcogliese, D. J. (2005). Parasites of the superorganism: Are they indicators of ecosystem health? *International Journal for Parasitology*, 35(7):705–716.

Marras, S., Killen, S. S., Lindström, J., McKenzie, D. J., Steffensen, J. F., and Domenici, P. (2015). Fish swimming in schools save energy regardless of their spatial position. *Behavioral Ecology and Sociobiology*, 69(2):219–226.

Mathis, A., Mamidanna, P., Cury, K. M., Abe, T., Murthy, V. N., Mathis, M. W., and Bethge, M. (2018). DeepLabCut: markerless pose estimation of user-defined body parts with deep learning. *Nature Neuroscience*, 21(9):1281–1289.

McElroy, E. J. and de Buron, I. (2014). Host Performance as a Target of Manipulation by Parasites: A Meta-Analysis. *Journal of Parasitology*, 100(4):399–410.

Mikheev, V. N., Pasternak, A. F., Taskinen, J., and Valtonen, T. E. (2013). Grouping facilitates avoidance of parasites by fish. *Parasites & Vectors*, 6(1):301.

Mohammed, R. S., King, S. D., Bentzen, P., Marcogliese, D., van Oosterhout, C., and Lighten, J. (2020). Parasite diversity and ecology in a model species, the guppy (*Poecilia reticulata*) in Trinidad. *Royal Society Open Science*, 7(1):191112.

Montgomery, J. C., Baker, C. F., and Carton, A. G. (1997). The lateral line can mediate rheotaxis in fish. *Nature*, 389(6654):960–963.

Moreno, K. R., Weinberg, M., Harten, L., Salinas Ramos, V. B., Herrera M., L. G., Czirják, G. A., and Yovel, Y. (2021). Sick bats stay home alone: fruit bats practice social distancing when faced with an immunological challenge. *Annals of the New York Academy of Sciences*, 1505(1):178–190.

Morgan, M. J. (1988). The influence of hunger, shoal size and predator presence on foraging in bluntnose minnows. *Animal Behaviour*, 36(5):1317–1322.

Nadler, L. E., Killen, S. S., Domenici, P., and McCormick, M. I. (2018). Role of water flow regime in the swimming behaviour and escape performance of a schooling fish. *Biology Open*, page bio.031997.

Nagy, M., Ákos, Z., Biro, D., and Vicsek, T. (2010). Hierarchical group dynamics in pigeon flocks. *Nature*, 464(7290):890–893. Number: 7290 Publisher: Nature Publishing Group.

Nicoletto, P. F. (1991). The Relationship between Male Ornamentation and Swimming Performance in the Guppy, Poecilia reticulata. *Behavioral Ecology and Sociobiology*, 28(5):365–370. Publisher: Springer.

Peig, J. and Green, A. J. (2009). New perspectives for estimating body condition from mass/length data: the scaled mass index as an alternative method. *Oikos*, 118(12):1883–1891. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1600-0706.2009.17643.x.

Plaut, I. (2001). Critical swimming speed: its ecological relevance. *Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology*, 131(1):41–50.

Poirotte, C., Massol, F., Herbert, A., Willaume, E., Bomo, P. M., Kappeler, P. M., and Charpentier, M. J. E. (2017). Mandrills use olfaction to socially avoid parasitized conspecifics. *Science Advances*, 3(4):e1601721.

Poulin, R. (1995). "Adaptive" changes in the behaviour of parasitized animals: A critical review. *International Journal for Parasitology*, 25(12):1371–1383.

Poulin, R. (2000). Manipulation of host behaviour by parasites: a weakening paradigm? *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 267(1445):787–792. Publisher: Royal Society.

Poulin, R. and Morand, S. (2000). The diversity of parasites. *The Quarterly Review of Biology*, 75(3):277–293.

Price, P. W. (1977). General Concepts on the Evolutionary Biology of Parasites. *Evolution*, 31(2):405.

Raffel, T. R., Martin, L. B., and Rohr, J. R. (2008). Parasites as predators: unifying natural enemy ecology. *Trends in Ecology & Evolution*, 23(11):610–618.

Rahn, A. K., Hammer, D. A., and Bakker, T. C. M. (2015). Experimental infection with the directly transmitted parasite Gyrodactylus influences shoaling behaviour in sticklebacks. *Animal Behaviour*, 107:253–261.

Reader, S. M., Kendal, J. R., and Laland, K. N. (2003). Social learning of foraging sites and escape routes in wild Trinidadian guppies. *Animal Behaviour*, 66(4):729–739.

Reebs, S. (2001). Influence of Body Size on Leadership in Shoals of Golden Shiners, Notemigonus Crysoleucas. *Behaviour*, 138(7):797–809. Publisher: Brill.

Reebs, S. G. (2000). Can a minority of informed leaders determine the foraging movements of a fish shoal? *Animal Behaviour*, 59(2):403–409.

Reynolds, M., Hockley, F. A., Wilson, C. A. M. E., and Cable, J. (2019). Assessing the effects of water flow rate on parasite transmission amongst a social host. *Hydrobiologia*, 830(1):201–212.

Richards, E. L., van Oosterhout, C., and Cable, J. (2010). Sex-Specific Differences in Shoaling Affect Parasite Transmission in Guppies. *PLoS ONE*, 5(10):e13285.

Ripperger, S. P., Stockmaier, S., and Carter, G. G. (2020). Tracking sickness effects on social encounters via continuous proximity sensing in wild vampire bats. *Behavioral Ecology*, 31(6):1296–1302.

Robinson, C. J. and Pitcher, T. J. (1989a). Hunger motivation as a promoter of different behaviours within a shoal of herring: selection for homogeneity in fish shoal? *Journal of Fish Biology*, 35(3):459–460.

Robinson, C. J. and Pitcher, T. J. (1989b). The influence of hunger and ration level on shoal density, polarization and swimming speed of herring, Clupea harengus L. *Journal of Fish Biology*, 34(4):631–633.

Roche, D. G., Binning, S. A., Bosiger, Y., Johansen, J. L., and Rummer, J. L. (2013a). Finding the best estimates of metabolic rates in a coral reef fish. *Journal of Experimental Biology*, page jeb.082925.

Roche, D. G., Taylor, M. K., Binning, S. A., Johansen, J. L., Domenici, P., and Steffensen, J. F. (2013b). Unsteady flow affects swimming energetics in a labriform fish (*Cymatogaster aggregata*). *Journal of Experimental Biology*, page jeb.085811.

Rosengaus, n., Jordan, n., Lefebvre, n., and Traniello, n. (1999). Pathogen alarm behavior in a termite: A new form of communication in social insects. *Die Naturwissenschaften*, 86(11):544–548.

Rosengaus, R., Traniello, J., and Bakker, T. (2022). Sociality and disease: behavioral perspectives in ecological and evolutionary immunology. *Behavioral Ecology and Sociobiology*, 76(7):98.

Rosenthal, S. B., Twomey, C. R., Hartnett, A. T., Wu, H. S., and Couzin, I. D. (2015). Revealing the hidden networks of interaction in mobile animal groups allows prediction of complex behavioral contagion. *Proceedings of the National Academy of Sciences*, 112(15):4690–4695.

Sankey, D. W. E., Shepard, E. L. C., Biro, D., and Portugal, S. J. (2019). Speed consensus and the 'Goldilocks principle' in flocking birds (Columba livia). *Animal Behaviour*, 157:105–119.

Seeley, T. D. (2010). *Honeybee Democracy*. Princeton University Press. Publication Title: Honeybee Democracy.

Simons, A. M. (2004). Many wrongs: the advantage of group navigation. *Trends in Ecology & Evolution*, 19(9):453–455. Publisher: Elsevier.

Smith, M. L., Davidson, J. D., Wild, B., Dormagen, D. M., Landgraf, T., and Couzin, I. D. (2021). The dominant axes of lifetime behavioral variation in honey bees. preprint, Animal Behavior and Cognition.

Spurrier, M., Boyce, M., and Manly, B. (1991). Effects of parasites on mate choice by captive sage grouse. *Bird-parasite interactions*, pages 389–398.

Sridhar, V. H. (2022). Inferring social influence in animal groups across multiple timescales. Manuscript submitted for publication.

Stephenson, J. F. (2019). Parasite-induced plasticity in host social behaviour depends on sex and susceptibility. *Biology Letters*, 15(11):20190557.

Stephenson, J. F. and Adelman, J. S. (2022). The behavior of infected hosts: Behavioral tolerance, behavioral resilience, and their implications for behavioral competence. In *Animal Behavior and Parasitism*. Oxford University Press.

Stephenson, J. F., Perkins, S. E., and Cable, J. (2018). Transmission risk predicts avoidance of infected conspecifics in Trinidadian guppies. *Journal of Animal Ecology*, 87(6):1525–1533. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/1365-2656.12885.

Stephenson, J. F., van Oosterhout, C., Mohammed, R. S., and Cable, J. (2015). Parasites of Trinidadian guppies: evidence for sex- and age-specific trait-mediated indirect effects of predators. *Ecology*, 96(2):489–498. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1890/14-0495.1.

Stephenson, J. F., Young, K. A., Fox, J., Jokela, J., Cable, J., and Perkins, S. E. (2017). Host heterogeneity affects both parasite transmission to and fitness on subsequent hosts. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 372(1719):20160093. Publisher: Royal Society.

Stockmaier, S., Bolnick, D. I., Page, R. A., Josic, D., and Carter, G. G. (2020). Immune-challenged vampire bats produce fewer contact calls. *Biology Letters*, 16(7):20200272. Publisher: Royal Society.

Stockmaier, S., Stroeymeyt, N., Shattuck, E. C., Hawley, D. M., Meyers, L. A., and Bolnick, D. I. (2021). Infectious diseases and social distancing in nature. *Science*, 371(6533):eabc8881.

Strandburg-Peshkin, A., Farine, D. R., Couzin, I. D., and Crofoot, M. C. (2015). Shared decision-making drives collective movement in wild baboons. *Science*, 348(6241):1358–1361.

Strandburg-Peshkin, A., Papageorgiou, D., Crofoot, M. C., and Farine, D. R. (2018). Inferring influence and leadership in moving animal groups. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 373(1746):20170006.

Stroeymeyt, N., Grasse, A. V., Crespi, A., Mersch, D. P., Cremer, S., and Keller, L. (2018). Social network plasticity decreases disease transmission in a eusocial insect. *Science*, 362(6417):941–945.

Sumpter, D. (2006). The principles of collective animal behaviour. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 361(1465):5–22.

Sumpter, D. J. and Pratt, S. C. (2009). Quorum responses and consensus decision making. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1518):743–753. Publisher: Royal Society.

Sumpter, D. J. T. (2010). *Collective Animal Behavior*. Princeton University Press.

Sumpter, D. J. T., Krause, J., James, R., Couzin, I. D., and Ward, A. J. W. (2008). Consensus Decision Making by Fish. *Current Biology*, 18(22):1773–1777.

Svendsen, J. C., Banet, A. I., Christensen, R. H. B., Steffensen, J. F., and Aarestrup, K. (2013). Effects of intraspecific variation in reproductive traits, pectoral fin use and burst swimming on metabolic rates and swimming performance: a study on the Trinidadian guppy (*Poecilia reticulata* Peters). *Journal of Experimental Biology*, page jeb.083089.

T. Russell, S., L. Kelley, J., A. Graves, J., and E. Magurran, A. (2004). Kin structure and shoal composition dynamics in the guppy, Poecilia reticulata. *Oikos*, 106(3):520–526. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.0030-1299.2004.12847.x.

Taylor, G. K., Nudds, R. L., and Thomas, A. L. R. (2003). Flying and swimming animals cruise at a Strouhal number tuned for high power efficiency. *Nature*, 425(6959):707–711. Number: 6959 Publisher: Nature Publishing Group.

Thomas, F., Poulin, R., and Brodeur, J. (2010). Host manipulation by parasites: a multidimensional phenomenon. *Oikos*, 119(8):1217–1223. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1600-0706.2009.18077.x.

Timi, J. T. and Poulin, R. (2020). Why ignoring parasites in fish ecology is a mistake. *International Journal for Parasitology*, 50(10-11):755–761.

Tunstrøm, K., Katz, Y., Ioannou, C. C., Huepe, C., Lutz, M. J., and Couzin, I. D. (2013). Collective States, Multistability and Transitional Behavior in Schooling Fish. *PLoS Computational Biology*, 9(2):e1002915.

Umberger, C. M., de Buron, I., Roumillat, W. A., and McElroy, E. J. (2013). Effects of a muscle-infecting parasitic nematode on the locomotor performance of their fish host. *Journal of Fish Biology*, 82(4):1250–1258. https://onlinelibrary.wiley.com/doi/pdf/10.1111/jfb.12061.

Van Oosterhout, C., Harris, P. D., and Cable, J. (2003). Marked variation in parasite resistance between two wild populations of the Trinidadian guppy, Poecilia reticulata (Pisces: Poeciliidae): Parasite resistance in the trinidadian guppy. *Biological Journal of the Linnean Society*, 79(4):645–651.

van Oosterhout, C., Mohammed, R., Hansen, H., Archard, G., McMullan, M., Weese, D., and Cable, J. (2007). Selection by parasites in spate conditions in wild Trinidadian guppies (Poecilia reticulata). *International Journal for Parasitology*, 37(7):805–812.

van Oosterhout, C., Potter, R., Wright, H., and Cable, J. (2008). Gyro-scope: An individual-based computer model to forecast gyrodactylid infections on fish hosts. *International Journal for Parasitology*, 38(5):541–548.

Walker, J. A., Alfaro, M. E., Noble, M. M., and Fulton, C. J. (2013). Body Fineness Ratio as a Predictor of Maximum Prolonged-Swimming Speed in Coral Reef Fishes. *PLoS ONE*, 8(10):e75422.

Walsman, J. C., Janecka, M. J., Clark, D. R., Kramp, R. D., Rovenolt, F., Patrick, R., Mohammed, R. S., Konczal, M., Cressler, C. E., and Stephenson, J. F. (2022). Shoaling guppies evade predation but have deadlier parasites. *Nature Ecology & Evolution*, 6(7):945–954. Number: 7 Publisher: Nature Publishing Group.

Walter, T. and Couzin, I. D. (2021). TRex, a fast multi-animal tracking system with markerless identification, and 2D estimation of posture and visual fields. *eLife*, 10:e64000.

Ward, A. J. W., Hoare, D. J., Couzin, I. D., Broom, M., and Krause, J. (2002). The Effects of Parasitism and Body Length on Positioning within Wild Fish Shoals. *Journal of Animal Ecology*, 71(1):10–14. Publisher: [Wiley, British Ecological Society].

Ward, A. J. W., Sumpter, D. J. T., Couzin, I. D., Hart, P. J. B., and Krause, J. (2008). Quorum decision-making facilitates information transfer in fish shoals. *Proceedings of the National Academy of Sciences of the United States of America*, 105(19):6948–6953.

Ward, J. H. (1963). Hierarchical Grouping to Optimize an Objective Function. *Journal of the American Statistical Association*, 58(301):236–244.

Webb, P. W. (1973). Kinematics of pectoral fin propulsion in Cymatogaster aggregata. *Journal of Experimental Biology*, page 14.

Wilson, A. D. M., Krause, S., James, R., Croft, D. P., Ramnarine, I. W., Borner, K. K., Clement, R. J. G., and Krause, J. (2014). Dynamic social networks in guppies (Poecilia reticulata). *Behavioral Ecology and Sociobiology*, 68(6):915–925. Publisher: Springer.

Xanthopoulos, P., Pardalos, P. M., and Trafalis, T. B. (2013). *Robust Data Mining*. SpringerBriefs in Optimization. Springer New York, New York, NY.

Östlund Nilsson, S., Curtis, L., Nilsson, G., and Grutter, A. (2005). Parasitic isopod Anilocra apogonae, a drag for the cardinal fish Cheilodipterus quinquelineatus. *Marine Ecology Progress Series*, 287:209–216.

Acknowledgments

It has been a colorful 72 months (except for the 6 times November).

First and foremost, I thank Iain Couzin. Thank you Iain for giving me the support and freedom (and time) to grow over the years. I learned critical and creative scientific thinking, to not give up, and to think big - with modesty. Every meeting with you is full of inspiration, and it motivates me and all the people in the environment you created in Konstanz. Thank you for making this possible. The Collective Behavior family is an amazing place for great science, friends and beers. And Thank you for being supportive to everyone also on non-work related matter, it makes the difference.

I thank Jacob Davidson, for patiently helping me figuring out my data analysis, my life, my thesis. Together, we have achieved a lot, organized TEDxKonstanz, finished my PhD, bought a van, gone to the mountains, and much more. I am very thankful for what you did for and with me. Let's soon go skiing to celebrate.

I thank Sandra Binning and Jessica Stephenson from my Thesis Advisory Committee for the countless zoom calls and the support over the last many years. It has been great to learn from you and I hope we will continue having great science discussions in the future.

Katja, Dago, Jenn: heroes. Without you we would all be in troubles. Thank you for being patient with all of us - questionable managers of ourselves. And thank you in particular for helping me, with the smallest form and more complicated work-life issues.

Thank you Mäggi and Daniel for the work you do for us and for the special treatment I always received by you; I often entered your offices to make noise and chat but you never kicked me out and instead, listened and supported me.

Thank you Margit for the tremendous work and the support I always received.

Francisca, thank you for making the IMPRS-QBEE students feeling part of a big family! And thank you for helping me for the most critical parts of PhD student: thesis submission, contracts, life.. thank you!

In the past years I have been surrounded by a lot of incredible people, and I am grateful to every single person that has contributed to making this time so exceptional. In alphabetical order:

Alex, talking to you is inspiring and I enjoyed learning from your approach to life and science; I am glad to have had the opportunity to meet for wine and talk, and spend time with your family.

Alexander, I love the care and kindness you use to make us all more educated and respectful towards any diversity.

Andrea, your kindness has taught me a lot. We started a lot of art projects together, and thank you for being so patient with me. I hope we will always find time for a wine and carving linoleum together.

August, thank you for all the snorkeling and the spontaneous night swimming. I enjoy the action with you.

Thank you Ben. I have spent the most incredible 6 years with you. Luckily, I was not grumpy with you at the IMPRS selection symposium, and I remember how hopeful I was to start this PhD together. Our friendship was a gem - probably ruby or sapphire. I have very warm memories of art, traveling, wine, bike rides, office, dinners, and much more. You are the opinion I'd like to have and the patience I lack. My family will miss you too. Moving to the other side of the world won't help you to get rid of me.

Cara, probably the best and only successful encounter on the SBB train. Thank you for all the artsy afternoons.

Ceci, thank you for all the times you didn't hesitate to come to my place and cook, bake a crostata with Escher's impossible figures on, oil paint on improbable copper surfaces. I hope we can soon bake a bitter orange crostata together. Strawberry is also okay.

Cami, Cate, Veri, una caressina a voi. Pezzi di cuore da tanti anni, e per molti di più.

Carlotta you are a flower. I have always felt home with you, Mati and Cami and had the best Covid lockdown I could ever imagine.

Dan B., thank you for teaching me electronics and starting the tradition of the beer hour.

And Thank you Daniel C. for keeping the beer hour tradition going, and for being such a caring person, I am grateful for the conversations we have had, about science and life.

Dan G. thank you for the music passion, I hope that after this thesis is over, we can enjoy more symphonic concerts together.

Eduardo, Emma, Saverio, Sophia I wish you arrived 5 years ago.

Thank you Edward and Nina for always feeding us and making us always feel welcome at your place. And thanks to the whole extended TND family.

Einat, thank you for the enthusiasm for the science and your support every time I wanted to record from a cockroach or a locust. I hope we'll manage to get recordings on insects again in the future.

Elena o Bonny, è circa dal lontano 2008 che mi sopporti, quindi grazie.

Erica, you have been essential for a long time, and gave me the confidence I needed for this.

Fritz you are art. I know that you can take me somewhere crappy, give me a beer, and make me feel as if I was most incredible place I could ever hope for.

Grace, thank you for the wines and the great chats. I love our deep conversations by the Bodensee.

Hemal, thank you for believing in the role of art in the life of scientists (and not only). I love the the kindness and calmness I feel around you.

Thank you single i Ian. Best office mate, and even if I tried to distract you from you PhD with all my love drama stories, you still managed to do an incredible science work and finish a decade before me. I missed you in Konstanz.

Inga, the truth is that with you I always had the great time, and wished we could do all the things like windsurfing, skiing, the back flips, the diving, the mountain biking and all the crazy actions. It's soon autumn and I am glad we now can do the Kürbissuppe and finally the Seidenmalerei.

Jacob S., with you I discovered that life can be much more. Flying, sailing, ski touring, building lamps, making pottery, building bikes, everything one might desire to do. Thank you for the deepest night talks and late night dart games. You are an invaluable friend. And thanks to Jette that makes us all dance.

Jacopo, I admire your powerful and intellectual approach to life and family. I look forward to discovering marine creatures at the isola di Bergeggi with you and Agata.

Jake, I would drink endless Negroni with you, and cover cupboards in gold leaf. Thank you for the countless help on science and other matters. I am glad we were part of this together.

Jan, thank you for all the breaks, coffees, roof-top conversations and the kicker. I am very glad we had many years together with fun, friends and work.

Jayme, it has been great to have you as a friend and colleague. Thank you for the great care you put in all the things, I am glad to have had you near me for both work and private sharing.

Jenna, we know it took us long but luckily we had a quick recovery and rapid improvement of our friendship. The next full moon is October 9th 2022.

Jessica Frosch you are „, awesome!!! Yay You are the most multifaceted human being I know. I started laughing with you underwater in the Seerhein swimming pool in 2017, and never stopped. I hope that wherever we will be in the future we will always do full moon bonfires, swim, hike, ski, cook, paint, party, cycle, and very importantly be yelled at by angry people. You are something else.

Joe, brief encounters but always full of scientific insight, thank you for the great chats.

Jolle, you helped me a lot in the beginning of my PhD, learning about sticklebacks and parasites, thank you.

Kaz, you are our sunflower. I admire your energy and the value it has for the people around you, in an outside work environment. Thank you for the most spontaneous laughs and fun, and the play catch!

Kiran, I am glad we got to share so much art and personal life. The inspirations I got from you will always motivate me to continue including art in my life.

Lucy, ho provato a trovarmi una nuova migliore amica ma ho fallito. Tu resti lì.

Luigi, your person is beyond simple wording. Thank you for believing in the fool person I am and to give me confidence in pursuing my sometimes unreasonable will. It's 'thanks' to you that I started my PhD, and I know you will keep inspiring me, even when the only thing I want to do is to cover buildings in gold leaf, and be pleased by it.

Luke, it has not been long but I had great science and art conversations with you. I want to do clay art with you and finally sit by the fire of your place with you and Clair after eating endless spaghetti with meatballs.

Marco, you were first the best supervisor I could hope for, and then the best friend. My time in Konstanz would have not been the same without you. I am glad we still always manage to have the best time together with a cocktail, good food, a collage, covid; anything really. And thanks to Leon and Camille, that are patient with you.

Marti, you have been essential for my past few year. I love the simplicity of our deep relationship, and the well-being I feel with you.

Matze, thank you for the endless conversations over beer on the Fürstenberg, and for the lack of agreement we often have. I can't count how many 'last' dart games we played in the middle of the night in the Cherisy, and it's probably true that you always win one game more than me.

Renaud you are an aesthetics genius, colors explode in your head and geometry appears, I want 100 portraits by you with the plotter. Now I know one important thing: art needs to be as big as possible. Thank you for all the most incredible art inspiration I got from you.

Serena, I admire your determination, you are a great friend and scientist, I hope we

can enjoy more fall/winter concerts at the Opernhaus in Zurich together soon.

Stefi, I always have a lot to learn from you, I would listen to you for hours (which is not hard to achieve). You always made me feel the 'support' in the most important moments, da buon maschio romano.

Tina and Andreas: you are my home and safe space. Our living together is precious.

Trillirillo, è un po' anche grazie a te se sono arrivata fin qui. Grazie.

Tuzzi, se un mio punto di riferimento da sempre. Grazie.

Tristan, You tried to scare me the first time we sat in front of each other at the selection symposium in Seewiesen: you failed. Even if Sicily is too laud for you, I know that you love Italians, even when they do not know (for many years) how to used Tristan's tracker and obsessively look for you.

Valentina, you have been my favorite conversation for over two years. I hope we will drink a negroni together, soon.

Vivek you make me feel warm and loved, and also manage to give me the confidence I often lack. Your kindness and sharpness will never stop to surprise me. Thank you for the great six years together, I hope there will be more. And I hope we'll finally go play squash together.

Yannick, you are the sweetest and badass scientist. You somehow never say NO to new projects but impressively enough you manage to do all. In a lot of ways I am grateful you are so busy so that we can have late night dinners at uni with Döner, wraps, pizzas and chats. I am glad I could count on you in the last most difficult weeks of the thesis writing, and anything before that.

Thanks to everyone else I shared lunches and beer hours with, to the whole Couzin Lab, the department of Collective Behaviour and our institute. And thanks to Katrin and Armin who brought a new light to Konstanz.

Thanks to the TEDxKonstanz team. It was an ambitious idea but I am very grateful for what we achieved. Thank you Carla, Jana, Kati, Mike, Lilith, Jennifer, Yannick, Katja, Barbara, Ian and Alexandra.

Thank you Mike, your knowledge in of art was always inspirational to me and I am grateful we also had a project together. Thank you for all the conversations and sharing. And thanks to Blair and Max, you are a wonderful family.

Thank you Lilith, the van worked fine and didn't make any noise, until it exploded.

Thanks to the Cherisy people and thanks to the Hüsli people, that made me discover the hippie and happy side of Konstanz.

Thank you Hans-Peter for taking me on boats and trusting my poor sailing skills.

Thanks to the Sella Group, un circolo di illuminati that manages to create the most improbable anagrams, excursions and kazoo concerts.

Thank you Senta for the greatest afternoons and the insightful conversations. It has been simply great.

Finally and most importantly, thanks to my family. Thanks to my mom and dad, who accepted me a long time ago and still support all my decisions - also when they take me far away from home to study fish. Thanks to my brother Francesco, the only person who will always know what is in my head. Thanks to Nicolò, a big source of happiness and Francesca, loving and caring. Thanks to my grandparents, Severina, Maria Grazia, Checco, Nico who see me little but always give me great affection and esteem (and food). Thanks to all my uncles and cousins with whom I grew up, and whose warm affection I always benefit from.

Thank you Albi cousins, soon we go for gotti ai preti. In particular thank you Vale, for believing in me and sharing valuable time and wine together.

Specific acknowledgments

I want thank all the animal caretakers, technicians, veterinaries and the animal welfare office that helped me at every step of the project. I want to thank Michael Mende, with the most difficult job to take care for all animal permits, lab space management, and coordination. I want to thank Daniel Zuniga and Inge Müller, doing a great job assuring animal welfare and good practices for animal experiments. I want to thank Basti and Benni to help keeping the fish in the best conditions possible despite the parasites - key for all the experiments. Thanks also to Alex Bruttel and Dominque Leo. Without any of the above mentioned people, my work would not have been possible. I also want to thank Jasminca Behrman-Godel and Myriam, for teaching me all I needed to know about fish and parasites, to give me space and the limnological institute and to help me in the earlier stages of the PhD.

Specific acknowledgments Chapter 1

Special thanks to Jayme Weglarski, who has given a huge contribution to the project, helping with animal holding, experiment preparation, room management, technical support, animal welfare and mental health. I also want to thank Christine Bauer, who has

helped with the initial phase of the experiment setup, with lab work running PCR on parasites, and in general with animal holding and animal caretaking.

Specific acknowledgments Chapter 2

I want to thank Johanna Stucke who helped with the experiments, animal caretaking, parasite counting. Running these experiments was a lot of work and it would have been very hard without the help of such a rigorous technician as you are.

Author contributions

Chapter 1

Angela Albi, Sandra Binning, Liang Li, Ellen Ye, Jake Graving, Jacob Davidson, Iain Couzin.

Angela Albi, Sandra Binning, Liang Li, Jacob Davidson and Iain Couzin conceived the project; Angela Albi performed the experiments and analyzed the data. Jacob Davidson, Liang Li and Sandra Binning helped the development of the method and data analysis. Jacob Davidson supervised data analysis and project outlining. Ellen Ye performed the tracking of the videos with DeepLabCut. Jake Graving helped with data analysis. Angela Albi led the writing of the manuscript; Sandra Binning, Liang Li, Jessica Stephenson, Jacob Davidson and Iain Couzin contributed critically to the drafts of the manuscript.

Chapter 2

Angela Albi, Jessica Stephenson, Sandra Binning, Tristan Walter, Vivek Sridhar, Jake Graving, Alex Jordan, Jacob Davidson, Iain Couzin.

Angela Albi and Iain Couzin conceived the project. Jessica Stephenson helped with project development, experimental design, and parasite culture. Alex Jordan helped with the experimental design. Angela Albi performed the experiments and analyzed the data. Tristan Walter helped with the tracking of the videos and developed software parts to aid the project. Jacob Davidson supervised data analysis and project outline. Jessica Stephenson, Sandra Binning, Vivek Sridhar, Jake Graving and Jacob Davidson and Iain Couzin helped with conceptualization of the results and analysis methods. Angela Albi led the writing of the manuscript; Jessica Stephenson, Sandra Binning, Jacob Davidson and Iain Couzin contributed critically to the drafts of the manuscript.

Chapter 3

Angela Albi, Vivek Sridhar, Máté Nagy, Jake Graving, Liang Li, Luke Costello, Jacob Davidson, Iain Couzin.

Angela Albi, Jacob Davidson, and Iain Couzin conceived the project. Angela Albi analyzed the data with supervision of Jacob Davidson. Vivek Sridhar, Máté Nagy, Jake Graving, Liang Li, Luke Costello and Iain Couzin helped with conceptualization of the analysis methods and results. Angela Albi led the writing of the manuscript. Jacob Davidson and Iain Couzin contributed critically to the drafts of the manuscript.