The Dynamics of Shoaling in Zebrafish

by

Noam Y. Miller

A thesis submitted in conformity with the requirements for the degree of Doctor of Philosophy

Graduate Department of Psychology
University of Toronto

© Copyright by Noam Miller 2010
The Dynamics of Shoaling in Zebrafish

Noam Miller

Doctor of Philosophy

Graduate Department of Psychology
University of Toronto

2010

Abstract

A wide array of species, from ants to humans, live or forage in groups. Shoaling – the formation of groups by fish – confers protection from predation and enhances foraging. However, little is known about the detailed characteristics or the dynamics of shoaling. Shoaling is a complex social interaction and a better understanding of its mechanisms and limitations would permit the study of natural and induced changes on social behavior generally in fish. Here, I present data on the shoaling characteristics of zebrafish (Danio rerio). Novel tracking techniques are used to extract detailed trajectories of all members of a free-swimming shoal of zebrafish. Multiple measures of shoaling – such as distributions of nearest neighbor distances, shoal polarizations, and speeds – are calculated, to better describe the subtleties of the behavior including, for the first time, the high resolution spatio-temporal dynamics of shoaling. In addition, a novel criterion is introduced to determine when and how individual fish or sub-groups leave the shoal.

Comparisons are presented between the shoaling characteristics of three populations of zebrafish (LFWT, SFWT, AB) and between days and hours of repeated exposure to the same testing environment, demonstrating the gradual effects of habituation on shoaling. In addition, the effects of manipulating the number of fish in the shoal, hunger levels, and predation threat are also examined, lending empirical support to ecological theories on the adaptive functions of the behavior. Finally, the data are compared to two leading theoretical models of shoaling and a
novel simulation approach is suggested. The data strongly suggest that various aspects of shoaling in zebrafish are constantly changing, complex, and flexible, representing a dynamic form of social cognition. The study of these characteristics sheds much-needed light on complex social interactions in this popular genetic model organism, which may eventually lead to a better understanding of social behaviors in other species, including our own.
# Table of Contents

## Chapter 1: Introduction

1.1 Natural history of zebrafish  
1.2 Behavioral ecology of shoaling  
1.3 Existing data on shoaling  
1.4 Mechanisms of shoaling  
1.5 Comparative studies  
1.6 Experimental measures  
1.7 Dynamics of shoaling  
1.8 Outline  

Figures  

## Chapter 2: Methods

2.1 General methods  
2.2 Tracking  
2.3 Analysis  

## Chapter 3: Coming and Going

3.1 Existing measures  
3.2 The current measure  
3.3 Results  
3.4 Discussion  

Tables and Figures  

## Chapter 4: Characterizing Zebrafish Shoaling

4.1 Distributions of basic measures  

Page

1  
3  
5  
7  
11  
12  
13  
15  
16  
17  
20  
22  
29  
35  
38  
42  
50  
53  
60  
61
4.2 Correlations between measures 64
4.3 Dynamics of shoaling measures 65
4.4 Discussion 67
Figures 73

**Chapter 5: Ecological Considerations** 80
5.1 Experiment 3: the effect of N 82
5.2 Experiment 4: the effect of food 86
5.3 Experiment 5: the effect of a predator 91
5.4 Discussion 95
Figures 98

**Chapter 6: Testing Models of Shoaling** 107
6.1 Methods for testing models against data 109
6.2 Testing existing models 110
6.3 A modest proposal 119
6.4 Discussion 125
Figures 127

**Chapter 7: Summary and Conclusions** 137
Figures 143

**References** 146

**Appendix A: Equations and Formulae** 159
**Appendix B: Additional Figures and Tables** 163
**Copyright Acknowledgements** 180
List of Tables

Table 3.1. Cross-correlations and results of PCA on excursion-related measures of shoaling.

Table B.1. Results of KS tests on excursion duration by day in Experiment 1.

Table B.2. Results of KS tests on excursion type by excursion duration in Experiment 1.

Table B.3. Results of KS tests on SGS by testing day for Experiment 1.

Table B.4. Results of KS tests on distance measures by day for Experiment 1.

Table B.5. Results of KS tests on speed by day for Experiment 1.

Table B.6. Results of KS tests on polarization by day for Experiment 1.

Table B.7. Results of KS tests on all unimodal measures by day for Experiment 3.

Table B.8. Results of KS tests on polarization distributions by day for Experiment 3.
List of Figures

Figure 1.1. Explanation of NND and IID.

Figure 3.1. Procedure for calculating the group membership criterion.

Figure 3.2. Mean number of excursions per session by day for each of the three strains in Experiment 1 (A) and by hour for Experiment 2 (B).

Figure 3.3. Density distributions of excursion duration by day for the SF shoals in Experiment 1 (A) and by hour for Experiment 2 (B).

Figure 3.4. Density distributions of the summed group size (SGS) by day for the SF shoals in Experiment 1 (A) and by hour for Experiment 2 (B).

Figure 3.5. Density distributions of position in the shoal before the beginning of an excursion, by strain for Experiment 1.

Figure 4.1. Density distributions of NND (A), IID (B), speed (C), and polarization (D) by day for Experiment 1.

Figure 4.2. Density distributions of NND (A), IID (B), speed (C), and polarization (D) by hour for Experiment 2.

Figure 4.3. Density distribution of IID including all fish (blue) and excluding fish on excursions (red) for Experiment 1.

Figure 4.4. Density distributions of speed (A) and polarization (B) by strain for Experiment 1.

Figure 4.5. Density distributions of bearing to nearest neighbor by strain for Experiment 1.
Figure 4.6. Modes of polarization distributions by day for all strains in Experiment 1, separated into schooling (open) and shoaling (solid) components.

Figure 5.1. The effects of shoal size on zebrafish shoaling.

Figure 5.2. Modes of polarization distribution components by shoal size.

Figure 5.3. The effects of food and hunger on zebrafish shoaling.

Figure 5.4. Density distributions of excursion duration (A), NND (B), speed (C), and polarization (D) for the final, probe, day for all groups of Experiment 4.

Figure 5.5. Modes of polarization distribution components for Experiment 4.

Figure 5.6. The effects of a predator image on zebrafish shoaling.

Figure 5.7. Density distributions of excursion duration (A), NND (B), speed (C), and polarization (D) for the final, probe, day for all groups of Experiment 5.

Figure 6.1. The zones in Couzin et al.’s (2002) model.

Figure 6.2. Viscido et al.’s (2004) AAR function.

Figure 6.3. Error distributions for Couzin et al.’s model for different values of the radius of zoop.

Figure 6.4. Error distributions by day for Couzin et al.’s model.

Figure 6.5. Error distributions by rank for Couzin et al.’s model.

Figure 6.6. Error distributions for Viscido et al.’s model for different values of $\delta_m$.

Figure 6.7. Error distributions by day for Viscido et al.’s model.
Figure 6.8. Error distributions by rank for Viscido et al.’s model.

Figure 6.9. Model suggestion measure distributions.

Figure 7.1. Spider-web graphs of all shoaling measures by strain for Experiment 1.

Figure 7.2. Spider-web graphs of all shoaling measures by day, for SF fish in Experiment 1.

Figure B.1. Experimental setup for Experiments 1, 2, 3, and 4.

Figure B.2. Experimental setup for Experiment 5.

Figure B.3. Density distributions of excursion durations by strain for Experiment 1.

Figure B.4. Density distributions of excursion durations by excursion type for Experiment 1.

Figure B.5. Density distributions of individual speed in the 1 sec before (blue) and the first sec of (red) an excursion in Experiment 1.

Figure B.6. Density distributions of NND by day for LF fish in Experiment 1.

Figure B.7. Density distributions of periodic oscillations in shoaling measures by strain for Experiment 1.
List of Appendices

Appendix A: Equations and Formulae.

Appendix B: Additional Figures and Tables.
Chapter 1: Introduction

Many species, from ants to humans, live or forage in groups (review in Sumpter, 2006). Amongst fishes, shoaling behavior is exhibited by almost half of all species at some point in their lives (Shaw, 1978) and has been proposed to confer multiple advantages in foraging and avoiding predation (see below; Krause & Ruxton, 2002). Shoaling and other group behaviors (herding, flocking) have been intensively studied at the ecological and evolutionary levels (Krause & Ruxton, 2002) but far less is known about the mechanisms underlying these behaviors. This imbalance is at least partially due to the difficulty in recording the detailed individual trajectories required for mechanistic analyses. Several research groups, including our own, have recently succeeded in gathering such data under carefully controlled conditions (e.g. Lukeman et al., 2010; Ballerini et al., 2008a, 2008b; Delcourt et al., 2009; Viscido et al., 2004). Using these detailed data, the studies presented here aim to characterize the shoaling behavior of zebrafish (Danio rerio).

Zebrafish have become a favorite behavior genetic model organism (Gerlai, 2003). They are easy to breed and house in large numbers and mature quickly (see below) and have thus been suggested as an efficient complement to other vertebrate models such as rats and mice (Guo, 2004). As such, and due to the wide array of genetic tools available for zebrafish, insights into their behavior can potentially be linked with relative ease to genetic and physiological mechanisms. Though findings on zebrafish behavior are accumulating rapidly (Miklosi & Andrew, 2006), a thorough characterization of zebrafish shoaling is still lacking.

1 A version of this chapter has been submitted for publication in Reviews in the Neurosciences and is currently under revision.
Behaving in a shoal presents several challenges. Shoals must cohere but not collapse on themselves; must evade predators and locate food yet not fragment for extended periods. How do fish solve these problems? Are shoaling mechanisms mostly hardwired and inflexible or can they be modified to a large extent by learning? Many species have been shown to modulate their shoaling based on environmental conditions (see below) but how flexible is zebrafish shoaling? Shoals navigate in their environment and can be led by a small percentage of informed individuals (e.g. Ward et al., 2008) but how do zebrafish leaders emerge (Conradt et al., 2009; see also Harcourt et al., 2009) and how are navigational conflicts resolved (Couzin et al., 2005; Harcourt et al., 2010)?

How fish solve these and related problems is of interest for several reasons. First, as demonstrated in the chapters that follow, shoaling is amongst the more complex behaviors exhibited by fish and understanding shoaling in detail may lead to more general insights about fish cognition. For instance, zebrafish are known to recognize individual conspecifics (Gerlach & Lysiak, 2006); could they then learn, for example, to shoal with certain ‘informed’ individuals and avoid ‘misguided’ ones, as has been attempted with chimpanzees (e.g. Povinelli et al., 1990)? Second, some of the mechanisms of shoaling may be generic to all group-forming animals and a detailed description of zebrafish shoaling may uncover basic rules that apply to any animal behaving in a group. Of course, the generality of such rules could only be established by examining other species at a similar level of detail. The conclusions from the present work (Chapter 7) suggest that a manageably small set of measures can be identified that reliably and comprehensively characterize an episode of shoaling and allow comparisons between different conditions or different species. This behavioral assay could also be used in combination with genetic and pharmacological manipulations to further explore the mechanisms of shoaling and social behavior in other group-forming species.
The present work initially presents several novel methods for examining shoaling in detail (Chapter 2) and then uses these to construct a thorough characterization of zebrafish shoaling (Chapters 3 and 4). The responsiveness of shoaling to environmental manipulations, and the ability of the measures selected to detect such changes, are examined in Chapter 5. Finally, the data are used to test existing models of shoaling and a suggestion is made for improving the models’ fit to the empirical data (Chapter 6).

Whilst shoaling is probably the most intensely studied form of collective motion, a large literature exists on herds of mammals and flocks of birds as well. To the extent that group forming in animals serves similar adaptive functions – enhancing foraging success and diluting predation – and to the extent that groups *qua* groups must share some elementary characteristics, these data add a great deal to our understanding of shoaling. However, unless a particular datum or, more often, a particular method is directly relevant, the mammal and bird literature has not been reviewed here. Swarms, groups of invertebrates, seem to follow different rules than do vertebrate groups (Sumpter, 2006) and the invertebrate literature is also not discussed here.

### 1.1 Natural history of zebrafish

The study of zebrafish behavior – which initially lagged far behind genetic, developmental, and pharmacological investigations – has increased dramatically over the past few years. Recently, researchers have examined genetic effects on behavior (Burgess & Granato, 2008), especially as they relate to disease models (Guo, 2004; Lieschke & Currie, 2007), simple behavioral responses primarily of larval zebrafish (Rihel et al., 2010; Chen et al., 2010), the learning capabilities of adult zebrafish (Williams et al., 2002; Bilotta et al., 2005; Colwill et al., 2005; Al-Imari & Gerlai, 2008; Pather & Gerlai, 2009), and the effects of various drugs on both learned (Levin & Chen, 2004; Levin et al., 2006; Eddins et al., 2009) and innate (Levin et al., 2007; Speedie &
Gerlai, 2008; Gerlai et al., 2008) behavior. Despite this surge in interest, much remains to be learned about zebrafish shoaling. The effects of environmental conditions, the effects of genetic differences between individuals or populations, and even the basic characteristics of ‘natural’ shoaling (i.e. absent any laboratory manipulation) all have yet to be described in any detail.

Zebrafish are small cyprinids, 2-4 cm long at maturity, found in warm slow-moving streams, ponds, and flooded rice paddies in southern and southeastern Asia (Engeszer et al., 2007b; Spence et al., 2008). They are found throughout the water column (Spence et al., 2006), usually in shallow areas close to banks (McClure et al., 2006). Zebrafish feed primarily on terrestrial insects that fall onto the surface of the water (McClure et al., 2006) and are preyed upon by a number of co-occurring fish (Engeszer et al., 2007b) – and possibly also avian – species.

Under laboratory conditions zebrafish spawn every few days throughout the year and can produce up to 200 eggs at one spawning. Mating most often takes place during the first hour after sunrise (Spence et al., 2006, 2008) during which time male zebrafish aggressively defend spawning territories (Spence & Smith, 2005). The fish remain active throughout the daylight hours (Hurd et al. 1998). Zebrafish display no parental care for their eggs or larvae. Larvae hatch at 72 hours post-fertilization and reach maturity at around 3 months in the lab but possibly as late as 6 months in the wild (Spence et al., 2008).

Male zebrafish form dominance hierarchies (Spence & Smith, 2006). Dominance may be independent of size (Spence & Smith, 2006) though female zebrafish do prefer larger males (Pyron, 2003). Dominance-related aggressive behavior may be limited to mating and spawning times (early morning), as male-male aggression and shoaling are mutually exclusive in related species (e.g. guppies; Magurran & Seghers, 1991).
1.2 Behavioral ecology of shoaling

The most widely accepted definition of a shoal is a group of fish that remain together for social reasons (i.e. not because of some attractive feature of their environment). A school, in contrast, is a shoal exhibiting synchronized, highly polarized collective motion (Pitcher & Parrish, 1993). This definition raises an obvious difficulty. Whilst the definition of a school depends on easily observable characteristics, that of the shoal does not. What are ‘social reasons’ and how would we determine the reason for any given grouping? What if there are both social and external reasons for aggregating? In addition, polarization varies smoothly both across groups and within a group across time. Is there a valid threshold value for schooling that clearly distinguishes it from shoaling, or do the two merge seamlessly into each other? Though the last of these questions may have a positive answer (see Chapter 4), in order to avoid confusion here all groups (of fish) are referred to as shoals, except where the distinction between shoaling and schooling is the subject of discussion.

The anti-predatory advantages of shoaling operate via several distinct mechanisms. A large shoal may be more likely to detect an oncoming predator – or detect it sooner – than an individual fish and, assuming this information is rapidly disseminated within the shoal, may be more likely to escape the attack. This is known as the ‘many-eyes’ hypothesis (Pitcher & Parrish, 1993) and should allow an individual in a large shoal to spend less time on vigilance and more time foraging than a singleton (Magurran & Pitcher, 1983). Visually-mediated social transmission of an alarm reaction has been demonstrated in zebrafish (Suboski et al., 1990), implying that they could benefit from the many-eyes effect. Even when attacked, assuming a predator cannot consume the entire shoal, a fish in a larger shoal is proportionally less likely to be targeted by the predator, an effect referred to as dilution (Krause & Ruxton, 2002). In addition, predators
attacking a rapidly dispersing shoal may experience ‘confusion’, being unable to select and follow a single target. Predator confusion may be enhanced by phenotypical homogeneity amongst the fish (Landeau & Terborgh, 1986).

The proposed foraging advantages of shoaling, particularly in the case of zebrafish, have not yet been as clearly demonstrated as the anti-predatory ones. Whilst zebrafish may be attracted to food sources by the actions of conspecifics (as demonstrated by Pitcher et al., 1982, for goldfish, Carassius auratus, and minnows, Phoxinus phoxinus), they may also compete for food and interfere with each other’s foraging (Wright et al., 2006b). Hamilton & Dill (2002) have shown that dominant zebrafish will monopolize a reliable stationary food source. Since zebrafish probably feed primarily on insects (McClure et al., 2006), they probably have little to gain from being alerted to the location of food as it is being consumed by a conspecific.

Shoaling may also facilitate locating potential mates and may provide hydrodynamic advantages (though this is debated; Pitcher & Parrish, 1993). Fish in larger shoals have proportionally more sources of social information, making it likely that social learning is enhanced in shoaling fish (Brown & Laland, 2003). Additionally, information transfer in shoals – which has lately been the subject of intensive research (e.g. Conradt & List, 2009, and other articles from the same special issue on ‘Group decision making in humans and animals’) – may be facilitated in larger shoals leading to a better utilization of available information on, for example, local predator and food distributions. Conversely, fish in a shoal may suffer greater exposure to parasites (but may also be better able to detect parasitized conspecifics) and predators may target large shoals more than small shoals or singletons, thus reducing the advantages conferred by dilution (Krause & Ruxton, 2002).
1.3 Existing data on shoaling

Shoals of different species vary in their average membership from less than 10 (e.g. guppies, *Poecilia reticulata*; Croft et al., 2003) to several hundreds of thousands (e.g. herring, *Harengula thrissina*; Parrish, 1992). Shoals are usually elongated along their direction of travel (Bumann et al., 1997) and display a higher density towards the front of the shoal (Partridge et al., 1983; Hemelrijk & Hildenbrandt, 2008). Fish maintain a distance of between 0.6 and 2 body lengths (BL) from their nearest neighbors (Partridge et al., 1983; Parrish et al., 2002) though spacing within the shoal varies with species, the size or speed of the shoal (Partridge, 1980; Partridge et al., 1980), and the age of the fish (van Olst & Hunter, 1970). Fish speeds and headings often correlate with those of their nearest neighbors (Partridge & Pitcher, 1980; Partridge et al., 1980) ensuring shoal cohesion. Most shoals disperse at night (e.g. Croft et al., 2003).

To the best of my knowledge, no-one has reported distributions of zebrafish shoal sizes, either in the wild or the lab. In the only published estimate I know of for zebrafish, Pritchard et al. (2001) suggested that zebrafish in the wild form shoals of 2-10 individuals but provided no data. Group size distributions in other species, including fish, have been found to follow truncated power laws (Bonabeau et al., 1999) and thus may not have a typical size (see also Krause & Ruxton, 2002). Preferred shoal size may also depend on environmental factors: guppy shoal sizes depend on the level of predation (Magurran & Seghers, 1991; but see Croft et al., 2003) with the majority of fish in low-predation populations found as singletons and shoals of about 20 individuals being most common in high-predation areas; minnows form larger shoals in the presence of a predator only in environments devoid of hiding places (Orpwood et al., 2008); banded killifish (*Fundulus diaphanus*) form smaller or larger shoals respectively as they are presented with food- or predator-related cues (Hoare et al., 2004).
Several researchers have attempted to document those characteristics of shoals that influence individuals’ preferences, usually by giving a test fish the option of joining one of two shoals that differ in some dimension of interest. The most common paradigm used involves dividing a tank into three sections separated by transparent barriers. The test fish is placed in the central compartment and different shoals (or video stimuli) are presented in the two side compartments. The amount of time the test fish spends in the vicinity of each side compartment is assumed to correlate with preference for one shoal over the other. Such studies have demonstrated that zebrafish prefer a shoal of conspecifics to an empty tank, even if the stimulus fish are of a different phenotype than the test individual (e.g. Sneckser et al., 2006) and different phenotypes of zebrafish prefer to shoal with conspecifics of a similar phenotype (Rosenthal and Ryan, 2005; Sneckser et al., 2010). These results suggest that zebrafish, though found in mixed species shoals in the wild (Spence et al., 2008), would display a preference for shoaling with conspecifics over heterospecifics. Other cyprinids, also found in mixed shoals, may only segregate into single-species shoals when there is an immediate threat of predation (Pitcher & Parrish, 1993).

Individuals of many shoaling species preferentially shoal with conspecifics of a similar size (Krause et al., 2000; Croft et al., 2003) and coloration (McRobert & Bradner, 1998) which has the effect of increasing the visual homogeneity of the shoal and thus potentially increasing predator confusion (Krause & Ruxton, 2002).

Zebrafish, like other species, prefer to join more numerous shoals, though this preference also depends on the activity level of the stimulus shoal (Pritchard et al., 2001) and the sex of the fish (male zebrafish prefer to shoal with females but females display no preference; Ruhl & McRobert, 2005). This preference may be due to the greater protection from predation that larger shoals are assumed to confer (Landeau & Terborgh, 1986).
Some fish species shoal only as juveniles, others throughout their life (Shaw, 1978). Zebrafish larvae begin to display a preference for conspecifics around the post-flexion stage (about 12 days post-fertilization; Engeszer et al., 2007a) and continue to shoal into adulthood. Engeszer et al. (2004) have shown that the preference for a particular phenotype in zebrafish depends on the fish’s rearing environment (see also McCann & Carlson, 1982) and is possibly learned during a critical period when the fish are juveniles (Engeszer et al., 2007a; but see Moretz et al., 2006).

In addition to preferring to join a particular shoal, fish may display a preference for certain positions within the shoal. The front or edges of a shoal are more dangerous than the interior, as they are more exposed to predators (Bumann et al., 1997) but may also provide better opportunities for foraging. Food-deprived roach (Rutilus rutilus) are more likely to take up positions in the front of a shoal than fed fish and frontal fish consume more food than other members of the shoal (Krause, 1993a). In shoals of chub (Semotilus atromaculatus), frontal fish are more likely to be attacked by a predator than their more central shoal-mates (Bumann et al., 1997) and minnows prefer positions in the center of a shoal over peripheral positions only after being frightened (by the application of alarm substance; Krause, 1993b).

Positional preferences may also reflect individual differences between fish within the same population. Roach found in the front of a shoal are more likely to retain their position than fish further back in the shoal (Krause, 1993a). Preferred shoal position, where it is a more-or-less constant trait, may correlate with other behavioral measures. Shoal leadership in sticklebacks correlates with boldness in non-social situations (Harcourt et al., 2009). More timid sticklebacks (Gasterosteus aculeatus; as determined by the time taken to recover from a simulated predator attack) are more likely to remain close to a stimulus shoal and less likely to take up frontal positions in a shoal (Ward et al., 2004).
In general, different positions within a shoal offer different opportunities and present different risks. Even assuming that each fish always occupies its preferred position, positions may not be constant for all fish and may vary with internal state (e.g. hunger) and environmental conditions (e.g. predation risk; Krause, 1994). Individuals may, in addition, be forced into certain positions in the shoal by the more strongly expressed preferences of their conspecifics.

Though shoaling has been shown to provide foraging benefits, it has also been suggested that foraging may be enhanced by gaining some distance from a shoal and that distances between individuals in shoals are maintained by a tension between the competing demands of safety and hunger (e.g. Krause & Ruxton, 2002). Food-deprived zebrafish shoals are less compact in the presence of food (Miller & Gerlai, 2007; but see Chapter 5 of the present work), which suggests that foraging success in zebrafish is enhanced by a loosening of the shoal and that the typical inter-individual spacing seen under conditions that do not encourage foraging is determined by other, most likely anti-predatory, considerations.

Hunger increases distances between juvenile Pollock (Theragra chalcogramma) and this effect is modulated by the presence of a predator (Sogard & Olla, 1997) and the distribution of available food (i.e. whether it is clumped or dispersed; Ryer & Olla, 1998). Similarly, in many other species, the presence of a predator or the threat of predation tend to increase shoaling tendency and shoal cohesion and hunger and/or the availability of food tend to decrease shoaling and loosen shoals (Krause & Ruxton, 2002). These competing effects are examined in detail in Chapter 5 of the current work.
1.4 Mechanisms of shoaling

In order to shoal, a fish must be aware of the locations (and possibly the speeds) of its shoalmates. In those species examined so far, shoaling is maintained through both vision and the lateral line (Partridge & Pitcher, 1980) and, in most species, shoals disband at night (e.g. Croft et al., 2003). Hunter (1969) showed that jack mackerel (*Trachurus symmetricus*) react to a change in direction by one shoal member within 0.5 sec when the group is in a highly polarized state. Nearest neighbors of the focal fish reacted first. Speeds and headings of shoals of minnows correlate best at a lag of around 0.3-0.6 sec (Partridge, 1980), though this value varies with the number of fish in the shoal. Partridge (1981) demonstrated that heading and speed in schools of saithe (*Pollachius virens*) correlate with those of at least their three nearest neighbors (but see Partridge & Pitcher, 1980). Larger shoals (more than about 10 individuals, for saithe) may consist of several intermixed subgroups that move somewhat independently of each other (Partridge, 1981).

Locating a shoal to join may be facilitated by odor cues. Bloom & Perlmutter (1977) showed that under some conditions zebrafish display a preference for water in which other zebrafish have been living. In juvenile zebrafish this preference extends to kin over non-kin and familiar kin over unfamiliar kin (Gerlach and Lysiak, 2006). Thus, zebrafish may preferentially shoal with kin or with familiar conspecifics, as has been demonstrated for guppies (Lachlan et al., 1998), though this effect may be modulated or even overridden by environmental factors (Morrell et al., 2007).

When injured, zebrafish skin (like that of other cyprinids) releases an alarm substance (Waldman, 1982). Alarm substance initially evokes an anti-predatory ‘escape’ response,
followed by a period of increased shoal density (relative to a pre-exposure baseline; Speedie & Gerlai, 2008), behaviors similar to those observed when a predator is detected (Miller & Gerlai, 2007). Similar results have been reported for guppies (Huizinga et al., 2009). Alarm substance has also been shown to increase erratic (possibly escape-related) movements in zebrafish in a dose-dependent manner (Speedie & Gerlai, 2008; Parra et al., 2009).

1.5 Comparative studies

Few studies have addressed comparative questions about shoaling. Obviously, many features of the shoaling of a large marine fish will be different from the shoaling of zebrafish. However, even within small freshwater species, there may be important and interesting variations in shoaling tendency and manner. For instance, Magurran and Pitcher (1983) showed that shoal size has different effects on foraging tactics in minnows and goldfish, which may be related to the generally greater shoaling tendency of the former species. As an excellent example of what can be done with detailed data, Partridge et al. (1980) compared the shoaling of cod (Gadus morhua), herring (Clupea harengus), and saithe and showed consistent species-specific differences in the regularity of spacing, density, and shape of the shoals.

The domestication of many species of fish may have led to important changes in their shoaling behaviors when compared to their wild-type progenitors (Wright et al., 2006b). For example, the constant food supply and greatly reduced level of predation in the lab may explain why a laboratory strain of zebrafish (AB) spends less time shoaling in some paradigms than wild-type derived fish (Wright et al., 2006a). Laboratory populations of zebrafish also spend more time close to the surface of the water than wild-type populations and are less fearful (Robison & Rowland, 2005).
1.6 Experimental measures

Empirical studies of group behavior have generally reported a number of standard measures of group cohesion. Common basic measures include the speed and bearing (heading) of an individual or of the group and the polarization of the group, which is the degree to which all members of the group are moving in the same direction.

The most common of all measures is the Nearest Neighbor Distance (NND; Figure 1.1), the distance between an individual and its nearest neighbor, usually reported in units of body lengths (BL). The NND is most often reported as the mean of the NNDs for all individuals measured (which is not always all members of the shoal). Clark and Evans (1954) have suggested a modified, normalized, NND that takes into account “the manner and degree to which the distribution of individuals… departs from that of a random distribution” (p. 446), but this measure has not been widely adopted. In addition to the first NND, it is possible to calculate the distance of a focal individual from its second nearest neighbor, or third, or $n$th.

The spatial distribution of $\text{NND}_n$ (the distance or heading from a focal fish to its $n$th nearest neighbor) has been used to ask how many of its neighbors affect the motion of an individual. Partridge & Pitcher (1980) have shown that velocities and headings of saithe correlate with those of their first two nearest neighbors at slow speeds but only with those of the first nearest neighbor at high speeds (see also Tien et al., 2004). Partridge et al. (1980) showed that bearing to the first three nearest neighbors in saithe and herring is non-random, implying that at least that many neighbors are influential. As a suggestive contrast, Ballerini et al. (2008a) recently showed that starlings ($\textit{Sturnus vulgaris}$) may attend to as many as 7 of their nearest neighbors. Partridge
et al. (1983) showed that the distribution of neighbors is also affected by the size of the shoal in bluefin tuna (*Thynnus thynnus*; see also Partridge, 1980).

Closely related to NND is the Inter-Individual Distance (IID; Figure 1.1), the mean distance between a focal individual and all other members of the group. IID is also often reported as a mean across all focal individuals (e.g. Warburton & Lazarus, 1991). We (Miller & Gerlai, 2008) recently demonstrated that the mean IID of zebrafish shoals oscillates with a characteristic period of about 5-15 sec (see also Chapter 4 of the present work).

Another popular measure is the shape of the shoal, though several different definitions of shape have been used in the literature. Early measures, such as Partridge et al.’s (1980), consisting simply of the ratio of width to length to depth of a shoal, had the advantage of being easy to grasp. Most current measures are based on the convex hull (the smallest convex polygon that contains all members of the shoal), but vary in their details. Partridge et al. (1983) describe several different shapes that shoals of tuna can take, and suggest that shape depends on the number of fish in the shoal. Shape may also vary with the speed of the shoal and be species-specific (Partridge et al., 1980). The shape of a shoal also correlates well with its polarization, more polarized shoals (schools) tending to be more elongated.

In addition to the measures discussed above, which attempt to characterize the structure of the entire shoal, it is possible to study the relative positioning of fish within a shoal (see above). For instance, a preference for being at the front of a foraging group may account for the increased frontal density observed in some shoals (Hemelrijk & Hildenbrandt, 2008). Position within a group is often defined in terms of whether a fish is peripheral or central. Peripheral fish are those that are on the convex hull of the group, and central fish are those that are not peripheral (Krause & Ruxton, 2002; but see Parrish et al., 2002, for a different definition). Black et al. (1992), for
example, have demonstrated that geese (*Branta leucopsis*) prefer the periphery of a grazing flock as it moves across a field but that peripheral birds expend more time and energy on vigilance than central birds.

### 1.7 Dynamics of shoaling

All of the common measures presented above may be considered ‘state variables’ of shoaling: they represent the condition of the shoal at a single point in time. Most authors to date (with the notable exception of Aoki, 1980) have reported the values of these measures as averages over several time points. Yet, as shown by Aoki (1980) for field gudgeon (*Gnathopogon elongatus*) and, more recently, by Miller & Gerlai (2008) for zebrafish, measures such as NND, IID, and polarization (Visicido et al., 2004) fluctuate widely during the course of shoaling. Members of a moving shoal constantly change their positions within the shoal, adjusting their distance from other members of the shoal, and even leave the shoal altogether (see Chapter 3 of the present work). To reduce all this complexity to a single average NND or IID score risks losing much of the fascinating structure inherent in the behavior. Some models of shoaling (see below) consider dynamic changes in the structure and/or cohesion of the shoal (though these are often ‘catastrophic’ phase changes; e.g. Couzin et al., 2002), but I am not aware of any empirical study except those by Aoki (1980) and Miller & Gerlai (2008) to measure and attempt to quantify the dynamics of shoaling. In the studies presented here, NND and IID are measured at frequent intervals (several times a second) and analyzed as time-series’ (Chatfield, 2002). These data should be of particular interest to modelers of shoaling as many current models, whilst reproducing the average characteristics of shoals quite well, often settle down to an unrealistically stable equilibrium solution (discussed, e.g., by Grégoire et al., 2003).
1.8 Outline

Examining the dynamics of shoaling required several novel statistical techniques. Time-series analysis methods are not well developed, especially in the life sciences, and several statistical methods had to be borrowed from other fields (engineering, economics) and sometimes adapted to the unique features of trajectory data. These techniques, and other general comments on experimental methods employed here, are presented in Chapter 2. In addition, determining the size of the group – on which several other measures depend – required developing a measure of group membership. This measure and the results of its application are presented in Chapter 3. Chapter 4 presents a characterization of zebrafish shoaling on multiple dimensions and data on the dynamics of shoaling at three different timescales, the first such data to be reported for any species, as far as I am aware. Chapter 5 presents the results of experiments designed to examine the effects of environmental manipulations (e.g. food-deprivation, predator presentation) on shoaling. Chapter 6 presents the results of comparing the data to existing models of shoaling and presents a novel approach that may better explain the dynamics of shoaling than existing heuristic models can. Finally, Chapter 7 summarizes the thesis.
Figure captions

Figure 1.1. Explanation of NND and IID. Each circle represents an individual (i.e. a fish). The Nearest Neighbor Distance (NND) of the focal (green) individual is the distance between it and the closest other individual, represented by the orange arrow numbered “1”. Distances to other individuals in the group are the second, third, or nth NND, represented by the other numbered arrows. An individual’s Inter-Individual Distance (IID) is the mean distance between the focal (green) individual and all other individuals; here, the mean of the lengths of all the arrows.
Figure 1.1
Chapter 2: Methods

The work presented in this dissertation involved the development of several novel techniques, both in data acquisition and analysis. In addition, the general methodology of all the studies presented in the following chapters was similar, and so is reviewed here.

Almost all previous studies of animal groups (with the exception of Aoki, 1980) have ignored the dynamics of collective behavior. Even when animals’ positions are measured at several closely-spaced time-points, the data are averaged together to give one score per session or segment of a session (e.g. Viscido et al., 2004). Either as a result of or as one of the causes of this lack of interest in dynamics, methods for the analysis of time-series data are not widely known in the life sciences. I have had to borrow, and often adapt, methods primarily from economics and engineering. In addition, there are few methods for dealing with non-independent measures (such as high resolution time-series data) without discarding much of the data (see below) and most such techniques do not focus on the unique problems of hypothesis testing. Techniques for breaking down the structure of time series’ – such as locating and describing oscillations – are also hard to find (economics texts tend to focus on prediction of fluctuations and engineering texts on their removal). As a result, several statistical tests described below are, as far as I am aware, novel adaptations of existing tests and more work needs to be done to firmly establish their validity.
2.1 General methods

2.1.1 Subjects. Subjects were all lab-reared zebrafish. Three different populations of zebrafish were used: Long-Fin Wild-Type (LF), Short-Fin Wild-Type (SF), and AB. These three populations were chosen for a variety of reasons. The AB strain has recently become the strain of choice for genetic, developmental, and, increasingly, behavioral studies. The LF and SF populations are both outbred. SF are phenotypically identical to ABs and represent the closest population to the original wild-type variety. LF zebrafish have extended fins and represent a different, but still outbred, population to the SF. In addition, it is possible that the extended fins of the LF affect their motor capabilities (Plaut, 2000) and this may have an effect on their shoaling. For simplicity, throughout this work all three populations are referred to as distinct strains.

Fish were all adults (3 months to 1 year old) and measured 3 to 4 cm in length (snout to fork of the tail). Groups of fish to be tested as a shoal were housed together for at least one week prior to the start of each experiment. Each group consisted of approximately equal numbers of males and females. During the experiment fish were housed in 40-litre tanks containing “system” water that was previously reverse osmosis purified and mixed with sea salt (‘Instant Ocean’ sea salt, Aquarium Systems Inc., OH) so that the conductivity of the water was between 900-1200 micro Siemens (576-768 TDS ppm). The water in the tanks was filtered (Aqueon PowerFilter 30, Franklin, WI), aerated, and maintained at a temperature of 26 ± 2 °C. Lights in the room in

---

2 The AB strain is a mostly inbred strain of zebrafish maintained in Oregon and is currently about 80-100 generations removed from its wild-type progenitors.
which the fish were kept went on at 7:00 h and off at 21:00 h. Fish were fed flake food (Tetramin Tropical Flakes, Tetra, USA) \textit{ad lib}, unless otherwise specified.

\textbf{2.1.2 Apparatus.} Fish were tested in a circular white plastic tank with a diameter of 91 cm, filled with system water to a depth of 10 cm. The temperature of the water in the tank was kept at 26 ± 2 °C. The tank was located in one corner of the room where the fish were housed. A moveable black room divider visually separated the tank from the rest of the room. The tank was lit by two fluorescent light fixtures placed at opposite sides of the tank, just above the lip of its wall. These ensured an even level of illumination in all parts of the tank and, being placed just above the level of the water, prevented glare (Figure B.1). Two different digital video cameras were used to film the experiments: in the earlier studies reported, a JVC Everio (model GZ-MG330) was mounted on a tripod attached to one of the walls of the room, with the camera lens 150 cm above the surface of the water; in the later studies a Sony Handycam (model HDR-XR-520) was attached to the ceiling with the lens of the camera 200 cm above the surface of the water. Both cameras were located such that the entire tank was visible in the frame of the video. The JVC camera produced video files at a resolution of 640 x 480 pixels, 10 fps; the Sony camera produced videos at a resolution of 1920 x 1080 pixels, 12 fps. Videos from both cameras were converted into AVI format using iSkySoft Video Converter (v 2.2.2.1).

\textbf{2.1.3 Procedure.} Fish were fed between 30 min and 1 hour prior to testing, except where the effects of food deprivation were being tested. All testing was performed between 10:00 and 14:00 h each day. Fish were gently netted from their home tank and transferred in a plastic beaker to the testing tank. They were then gently poured into the center of the tank. Fish were filmed for the duration of the session, usually 30 min, and then were gently netted into the beaker and returned to their home tank. The water in the testing tank was replaced after each two groups
had been tested to control for possible odor cues left by previous groups. In all experiments, all conditions were run concurrently and group order was semi-random.

2.2 Tracking

Fish were tracked from the video files using a custom application. Several authors have published results of experiments using custom automated tracking software but no such application has ever been made publicly available, to the best of my knowledge. Thus, I designed an automated tracking program, described in detail below. A brief review of the various similar software applications mentioned in the literature follows.

Several general principles of video tracking, conserved across almost all current applications, may be enumerated. Almost without exception, target locations are identified by subtracting the image of interest (a single frame of video) from a reference image of the testing enclosure without the targets. Reference images are either frames of video filmed before the subjects are introduced into the arena, or are created digitally post hoc by erasing the targets using image-editing software (e.g. Tien et al., 2004) or averaging multiple images together (which has the effect of smoothing out the mobile targets and retaining the stationary background; e.g. Viscido et al., 2004). Any pixels in which the difference between the reference and target images exceeds some threshold are flagged as part of a target.

Once the locations of the targets in a single frame have been identified, most current applications go on to attempt to reconstruct the trajectory of each target through successive frames (i.e. through time). The computation required may be thought of as mapping each target in frame T to the corresponding target in frame T-1. Though the details vary (and are rarely reported) most systems use a form of regression in which the mapping giving the smallest summed spatial
deviation of targets is assumed to be the correct one. This requires that the density (in time) of tracked frames be quite high so that the targets do not move very far between successive frames.

Finally, the most computationally difficult problem faced by all video tracking systems arises when targets occlude. Occlusion occurs when two or more targets move under or so close to each other that they appear as one target. Although it is generally accepted as a valid compromise to assign all participating targets the same coordinates during the occlusion, the difficulty lies in deciding, after separation, which of the targets corresponds to each target before the occlusion. Experimenters have sought to reduce the severity of this problem by testing fish in very shallow water (often as little as 2 cm; e.g. Becco et al., 2006) or by tracking their targets in all 3 dimensions using multiple cameras (e.g. Viscido et al., 2004; Ballerini et al., 2008b), a solution that creates its own problems. No consistently successful algorithm for solving the problem of occlusion has yet been devised, to the best of my knowledge, and most current systems permit the user to intervene and manually correct errors or resolve ambiguities.

2.2.1 Manual tracking systems. The earliest computerized tracking systems, such as the GALATEA system (Potel & Wasserzug, 1981), were manual, requiring the user to identify all the targets from an image or frame of video. One such application, still in use in our lab, was presented by Miller & Gerlai (2007). This application maximized the efficiency of coding by automating everything but the identification of the fish in the image – computationally the most difficult process. The user identified the locations of the fish by clicking with the mouse on top of the frozen video frame. The program automatically presented the next frame of video to be coded once a predetermined number of clicks had been registered and output to file not only the coordinates of each target but also a range of pre-selected measures such as NND, IID, area (of the convex hull of the school), and distance from any predefined point or line (e.g. the distance
from one edge of the tank). Despite the relatively high degree of automation of this application – in comparison to similar systems – the amount of user-hours required to code a reasonable number of video frames is still prohibitive. As a result, manual systems have most commonly been used to track no more than one frame per second (often significantly less), making any attempt at trajectory reconstruction, at least in fish, impossible. In addition, when coding manually, coder effects (differences between coders) may contaminate the data, though this effect has been found to be minimal (Miller & Gerlai, 2007).

2.2.2 Automated tracking systems. Applications for automated tracking of single targets have been commercially available for some time. The most widely known is Ethovision (Noldus Information Technology). Despite many complex options and add-ons, even the latest version of Ethovision (XT) and its commercially-available competitors cannot track more than one target per arena (unless the targets are of different colors).

A few research groups have now developed automated tracking systems for multiple identical targets. However, details of these applications have rarely been published nor have any of the applications been made publicly accessible. Thus, it is often difficult to compare different applications or to evaluate their precision.

Viscido et al. (2004; see also Grünbaum et al., 2005) used two video cameras and a set of macros designed by the NIH (known as NIH Image) to extract the 3D positions of shoals of either 4 or 8 giant danios (Danio aequipinnatus). Using 3D coordinates, which requires the use of at least 2 synchronized cameras, is an attempt to solve the problem of occlusion. Viscido et al. (2004) then used their own custom application called Tracker3D to reconstruct the trajectories of the fish. Tracker3D uses the minimal summed deviation between frames (see above) to determine target identity. Fragmentary trajectories were then ‘stitched’ together manually by the user. Trajectories
created from each camera (each in 2D) were then combined by assuming that trajectories that passed closest to each other belonged to the same target.

Becco et al. (2006; see also Delcourt et al., 2006, 2009) tracked up to 80 Nile Tilapias (Oreochromis niloticus L.) using their custom tracker. This system, in addition to the standard regression used to identify targets in consecutive frames, employs a trajectory prediction algorithm. Trajectory prediction algorithms use the previous heading of a target together with a known maximal turning rate to limit the region within which the application searches for the target (Delcourt et al., 2009). Specifically, targets must lie within a cone with its tip at their previous location and of twice the width of their maximal turning angle. By limiting the regression algorithm to solutions that respect these restrictions, target identification may be improved.

Recently, members of the STARFLAG project (Ballerini et al., 2008a, 2008b) have reported tracking large flocks of starlings in flight (up to 2600 birds in some flocks) using novel algorithms to identify the 3D locations of a large fraction (but not all) of the birds in the flock. The STARFLAG project is also unique in that the researchers have published detailed descriptions of their algorithms and apparatus (Cavagna et al., 2008). However, the STARFLAG analysis was primarily concerned with the momentary configuration of the flocks and no attempt was made to compute trajectories of individual birds (see also Cavagna et al., 2010).

Thus, no pre-existing application was available to track the data for the current work. As a result, a custom application was designed, described below. A more technical review of variations in current tracking applications is given by Delcourt et al. (2009).
2.2.3 Multiple Target Tracker. The application that I designed for tracking zebrafish is called Multiple Target Tracker (MTT). The program has gone through several versions since it was first designed; the current (and final) version, used for tracking all of the data presented in the current work, is described here. MTT is divided into two separate applications: Tracker and TrajMaker, each of which deals with one of the two main stages of tracking: identifying the targets and reconstructing their trajectories.

Tracker identifies the positions of a predetermined number of targets in a sequence of video frames and outputs the coordinates of each target in each frame to a text file. The process consists of three steps: identifying target pixels, clumping pixels into targets, and splitting targets that are too large.

The first step of the process involves subtracting the target frame from the reference image. As the videos are in color, each pixel has 3 values: red, green, and blue. The three values are summed for each pixel of the reference image and for the corresponding pixel in the target frame and the difference between the values computed. A user-defined threshold is used to identify pixels that may belong to a target.

The second step involves ‘clumping’ the above-threshold pixels into targets. The computational challenge here is to decide which pixels belong to the same target and which to an adjacent target. Each clump is seeded from a single pixel. The clump then ‘grows’ outwards by recruiting other above-threshold pixels that are close enough to the preceding pixel to be considered part of the same target. A user-defined parameter, Tolerance, determines how much space (in pixels) between above-threshold pixels can be bridged by the clumping algorithm. Thus, a Tolerance value of 2 means that there can be no more than 2 intervening pixels between two above-threshold pixels that are to be considered part of the same target. The growth process is iterated
until the target no longer grows with repeated iterations. Then, the next as-yet unassigned pixel is used to seed the next clump and the process repeats until all above-threshold pixels have been assigned to a clump.

The third and final step of the tracking process consists of identifying clumps that are too large or too small to be a single fish and modifying them accordingly. The minimal (MinT) and maximal (MaxT) number of pixels that a true target can occupy are determined by the user (similar parameters are used by Noldus’s Ethovision). Clumps consisting of fewer pixels than MinT are removed from further consideration. Clumps consisting of more than MaxT pixels are assumed to have resulted from partial occlusion (two or more fish occupying the same clump) and are therefore successively split until each clump contains fewer than MaxT pixels. Each over-large clump is split along its shortest axis: the shortest line that passes through the centroid of the clump is identified and pixels on either side of the line are assigned to different daughter clumps. If either daughter clump is still made up of more than MaxT pixels, it is split again.

After all three steps have been completed, the mean location of all the pixels constituting a particular clump is taken as the location of a target in the current frame. Tracker is blind to the expected number of targets in the video.

TrajMaker, the second application in the process, reads in the files created by Tracker and reconstructs the trajectories of individual fish. Computationally, TrajMaker assigns an identity to each target in each frame and attempts to ensure that identities are consistent across frames. TrajMaker uses a similar regression method to other tracking applications (e.g. Viscido et al., 2004), assigning identities based on the minimal summed deviation of target locations between successive frames. For each frame, the summed distance moved by all targets is calculated for every possible mapping of target identities and the minimal solution is selected. No trajectory
prediction algorithm is used as they are of dubious advantage when applied to zebrafish, which can turn very quickly.

_TrajMaker_ employs one more simple technique to ensure that targets are not misidentified during tracking. As fish may sometimes be missed by _Tracker_ (due, for example, to reflected light off the surface of the water reducing the image to below the image subtraction threshold), _TrajMaker_ has a user-defined movement threshold. If the closest match of a target is further than the threshold distance (in other words, the target ‘moved’ more than the threshold distance in one frame), it is assumed that the target was not tracked in the current frame. The best mapping of target identities is recalculated, excluding the missing target. The closest of any remaining unassigned targets after the mapping is complete is tentatively presumed to be the missing target. The program pauses and the user is required to manually identify the location of the missing target (the last known position of which is indicated). The movement threshold is set with the maximal speed of zebrafish in mind.

The main challenge for any tracking system is dealing with occlusions of targets. _TrajMaker_ employs either of two different methods to resolve the ambiguity of occlusions: one automated and one manual. If the manual option is selected, _TrajMaker_ allows the user to resolve all occlusions (as in, e.g., Viscido et al., 2004). Suggested trajectories for the relevant fish for the duration of the occlusion are overlaid on the video, which can be played forwards or backwards at any speed. The user selects the correct trajectory from all the possible combinations and the program continues tracking automatically until the next occlusion occurs. It is important to note that occlusions often involve more than two fish (sometimes all of the fish) and even an experienced observer is sometimes unable to unambiguously assign target identities. Under the automated method of solving occlusions, _TrajMaker_ calculates the angle of turn of each target.
under each possible combination of target identities and selects the option that minimizes the summed total turn angle. This is somewhat similar to the trajectory prediction algorithms described above. Delcourt et al. (2009) have recently described an application that uses the same method. The algorithm assumes that fish more-or-less retain their heading for the duration of an occlusion, an assumption which may well be false. Nonetheless, no better method has been suggested in the literature.

Whilst TrajMaker outputs complete trajectories for each fish in the shoal, it is important to keep in mind that some trajectories may be incorrect, particularly at points of occlusion. An informal comparison of the manual and automatic methods employed by TrajMaker leads to the conclusion that they do not differ by much (a similar analysis was carried out by Delcourt et al., 2009) and all the experiments reported here were coded using the automated method.

In conclusion, MTT is a mostly automated tracking system for multiple identical targets. The user is still required to resolve some ambiguities and to locate any fish that the tracking system misses. Many of the operations MTT carries out are computationally intensive and the program does not work very quickly. Tracker, the first application, requires about 10 seconds to track one frame of video (running on an 2.7 GHz Intel Dual-Core PC), or about 8 hours for 5 minutes of video (at 12 fps). TrajMaker, the second application, takes about 30 min to code the same amount of data (TrajMaker is much faster as no actual image processing is required). Only the latter application requires the participation of the user after the initial setup.

2.3 Analysis

The trajectory data produced by MTT were imported into Mathematica (v 4.0, Wolfram Technologies). A custom notebook then analyzed the data, according to the demands of each
experiment. The following sections describe the basic analyses common to all the studies and some of the fundamental statistical issues that had to be dealt with. Precise formulae for all measures are given in Appendix A. A significance level of 0.01 was used for all tests.

**2.3.1 Statistical concerns.** The trajectory data collected here are difficult to analyze statistically, as adjacent – and often also more distant – data points are non-independent. The Kolmogorov-Smirnov (KS) test – which compares two density distributions to each other – was modified to permit comparisons of distributions of non-independent measures. A common method of analyzing non-independent data is to resample the data at longer intervals, thus reducing the non-independence of adjacent points. The resampling interval chosen is usually the first zero-crossing of the autocorrelation function of the data, i.e. the interval at which consecutive values are maximally uncorrelated (for an example using behavioral data see Inman, 1990). This method is less than ideal, particularly for the data examined here. Autocorrelations in time-series’ represent two independent sources of correlation: one resulting from the proximity in time of consecutive measurements; the second from behavioral consistency in the subject. The former is noise; the latter data. Additionally, the measures examined here have extremely long-term autocorrelations, far longer than would be expected simply from the over-sampling of the data (autocorrelations for IIDs of individual fish, for example, were found to remain well above 0 for almost a quarter the length of a dataset, well over a minute). Thus, using the standard method here would result in such a low effective N that all comparisons would be rendered insignificant, even in cases where the distributions are ‘obviously’ vastly different. Thus, as a compromise, the effective N for each KS test was divided by the number of frames coded per second (usually 12). In other words, the data were effectively resampled to once per second. This value seems reasonable given the speed of movement of zebrafish, yet is not so large as to render the comparisons between distributions meaningless. Despite being completely arbitrary, the same value was used consistently for all
distribution comparisons (i.e. I did not titrate the resampling coefficient for each dataset to generate significant results on a whim). This procedure is referred to below as the ‘modified’ KS test.

2.3.2 Data preparation. Before any measures of shoaling were calculated, the data were smoothed using a weighted moving average with a window width of 0.5 sec (Chatfield, 2002). In addition, all the coordinates were rescaled – using the known width of the arena – so that all values were in centimeters. The origin of the axes was placed in the lower left-hand corner such that all the positions had positive coordinates, for simplicity.

Initially, several basic descriptive statistics (NND, IID, speed, and polarization) were calculated for every dataset. Each measure describes a univariate time-series, tracked at either 10 or 12 frames per sec. Each measure was interrogated at three levels, described in greater detail below: the distribution of the measure, its dynamics, and correlations between different measures (see Appendix A for definitions of each measure).

2.3.3 Measure distribution. The density distribution of each measure was determined using a kernel density estimator (KDE, essentially a smoothed histogram; see Appendix A). Distributions were calculated for the mean NND (per frame), the mean IID, the mean speed, and the polarization of the shoal. Initially, several other measures were also examined (such as the speed of the shoal centroid and individual NNDs and IIDs) but correlation analyses revealed that these extra measures provided little additional information and presented no new phenomena and, in the interests of brevity and clarity, they were dropped.

One advantage of studying the distribution of a measure (rather than the mean), other than the greater amount of information it conveys, is that distributions are additive. Thus, distributions
from different groups under the same experimental condition were averaged together and compared to each other using the modified KS test.

2.3.4 Measure dynamics. In addition to the distributions of measures, the time-series of each measure was examined directly. We have previously demonstrated that the mean IID of zebrafish shoals oscillates with a distinctive period (Miller & Gerlai, 2008). Aoki (1980) showed similar oscillations in the NND of field gudgeon shoals. Viscido et al. (2004) showed that the polarization of shoals of giant danios (Danio aequipinnatus) oscillates as well (see their Figure 2). Thus, none of these measures is constant across time. We therefore examined the time-series of each measure for both periodic components and long-term linear trends (across repeated exposures to the testing environment).

Periodic oscillations in the time-series’ were examined using the discrete Fourier transform of the data. The fast Fourier transform (FFT) algorithm was used to calculate the periodogram of each time-series. The periodogram is a plot of the period of each oscillation (the inverse of the frequency) against the power of each period (i.e. it displays how much of the variance in the series is accounted for at each period). Peaks in the periodogram represent potentially significant oscillation periods in the time-series. In order to determine which of the periodogram peaks were significant, we used the Lomb-Scargle test (Hernandez, 1999; Frescura et al., 2007; see also Appendix A). For each measure, the distribution of significant peaks (pooled from all sessions under the same experimental condition) was constructed. Distributions were compared to each other using a KS-test (unmodified in this case, as the peaks of a periodogram are independent of

---

A slightly different method of testing for oscillations was described by Miller & Gerlai (2008). However, the method described here is, in my opinion, more robust. All the data from the experiments described here were subjected to the newer analysis.
each other). This method, to the best of my knowledge, has not previously been applied to animal data (but see Miller & Gerlai, 2008).

2.3.5 Measure interactions. NND and IID are related to each other in several obvious ways. For example, if a focal fish’s NND is large, its IID will most likely also be large (as all the other fish are even further away than the nearest neighbor). Speed and polarization have also been shown to correlate (e.g. Viscido et al., 2004), implying that shoals are either fast and polarized or slow and disorganized. Thus, we examined the correlations between different measures as a way to determine the parameter space inhabited by zebrafish shoaling.

2.3.6 Bearing to nearest neighbor. Several authors have examined not only the distance of a focal fish from its nearest neighbor(s) but also the relative direction of those neighbors. The measure most often used, also adopted here, is the bearing to the nearest neighbor, which is defined as the angle between the heading of a fish and the bearing from it to its nearest neighbor (or its second, third, etc. nearest neighbor). This measure has been used, amongst other things, to estimate how many of its nearest neighbors a fish takes into account when positioning itself (e.g. Partridge, 1981). It is assumed that the bearing to the nearest neighbor corresponds directly to the preferred positioning of each individual fish. Whilst this may hold for large shoals, it remains to be seen (and has not yet been demonstrated) that it is equally true of small shoals, as used here. Nonetheless, reporting bearing to nearest neighbor has become de rigeur in studies of shoaling.

2.3.7 Individual differences. The data collected in the studies reported here allow us to track the identity of individual fish for the duration of a dataset (5 min) with a high degree of certainty. Whilst this is not quite as good as also keeping track of individual identities between sessions, it nonetheless allows an analysis of individual differences between fish in the same shoal. We examined one particular aspect of individual differences between fish: positional preferences
within the shoal (the existing data and theories on positional preferences are reviewed in Chapter 1). Position was coded as rank in the shoal. In each frame, all the fish were ranked according to how close they were to the front of the shoal (i.e. 1 = leading the shoal; 8 = at the back of the shoal). Position was determined relative to the direction of movement of the shoal, which was taken as the direction of the summed vector of the headings of all fish. Ranking distributions were compared to a uniform distribution using the modified KS test.

Finally, in addition to all the analyses described above, each file was also analyzed for the effects of fish leaving and rejoining the group. This analysis and its results are described in detail in the following chapter.

2.3.8 Conclusion. The current analysis is intended to provide a more comprehensive and nuanced picture of zebrafish shoaling than has previously been published. This is achieved by reporting and analyzing a number of different measures relating to the shoal and its constituents. These measures, though not entirely independent of each other, may be thought of as different dimensions defining the conceptual space within which shoaling occurs (see Fonio et al., 2005, for a similar approach). In addition, apart from any direct linear correlations between different measures, the effects of categorizing shoaling episodes based on some of the measures on patterns in the other measures were examined, such as reexamining the effects of polarization when fish that are away from the group (Chapter 3) are excluded from the analysis. The following chapters examine in what ways habituation (Chapters 3 and 4) and various environmental manipulations (Chapter 5) affect the characteristics of shoaling, as uncovered by changes in the above measures.
Chapter 3: Coming and Going

Zebrafish, though considered obligate schoolers, do not spend all their time in shoals (as herring, for instance, probably do). Delaney et al. (2002) placed zebrafish in a large enclosure that contained artificial plants. After the first three hours, spent in shoals, zebrafish dispersed and spent their time either singly or in pairs in the vicinity of the plants, presumably since hiding amongst plants provides similar anti-predatory benefits to shoaling. However, before a more detailed examination of the formation, dissolution, and flexibility of zebrafish shoals is possible it is necessary to quantitatively define shoal membership. This chapter presents a novel criterion for shoal membership that more accurately represents the dynamics of the shoal than existing measures.

3.1 Existing measures

There are very few criteria in the shoaling (and flocking and herding) literature for determining when or whether a fish is a member of a shoal. Most pelagic shoals rarely break apart and form sharp edges and can thus be delimited by eye with relative ease. Many authors have thus simply assumed that all the fish in a testing tank or in a region of study are part of the same shoal (e.g. Partridge et al., 1980) or have decided by eye, informally, where the limits of the shoal lie (e.g. Krause, 1993a).

3.1.1 Elective Group Size. Pitcher et al. (1983) introduced the first popular measure of shoal membership, the Elective Group Size (EGS). Following the rationale that only fish that are in some form of communication with each other should be considered part of the same shoal, the authors set a limit of 4 body lengths (BL) as the maximal distance between fish in the same shoal. Of course, by this criterion, the farthest two fish in a shoal can be many hundreds of
meters apart – as fish in large pelagic shoals often are – as long as each individual’s nearest neighbor is no more than 4 BL away. Mean EGS distributions have been reported for a variety of species (e.g. Pitcher & Parrish, 1993).

The EGS was widely adopted, though some authors substituted alternate, equally arbitrary, distance thresholds (e.g. Viscido et al., 2004, use 5 BL; Budaev, 1997, uses 7 BL). The main advantages of the EGS are its intuitive simplicity, the ease with which it can be measured, and the possibility of using it for comparisons between species (though this has not often been done). However, the simplicity of the measure is also its greatest drawback. No study that I am aware of has examined how shoal size distributions vary under different threshold values for the EGS, nor whether Pitcher et al.’s (1983) often-cited 4 BL threshold has any empirical basis as a limiting distance for communication. Nor has the question whether different species might have different thresholds (due, for example, to different perceptual capabilities) been addressed. In addition, the EGS ignores the distances between any but nearest neighbors.

3.1.2 IID-based measure. Miller & Gerlai (2008) suggested a group membership criterion that takes into account the spacing of all the fish within the shoal. In addition, this measure was the first (other than Aoki’s; see below) not to impose a fixed distance criterion but attempt to derive the threshold from the positional data. Our approach stems from the conviction that group membership at any given moment depends on the overall spatial distribution and density of the group at that moment. As these values are not constant, a constant distance threshold cannot accurately capture shoal membership. Intuitively (though our intuitions in this regard may not provide the best guidance), a fish that is 4 BL away from a group of conspecifics that are all 0.5 BL from each other is not ‘part of the shoal’ in the same sense as a fish that is 4 BL away from a group of conspecifics that are all 3 BL from each other. Our criterion, which attempted to take
this difference into account, is based on the IID of each fish. The IID is a better tool for the task at hand than the NND as it is affected by the distances between all the fish. However, this also means that the IID of a larger (more numerous) group will tend to be larger than that of a smaller group. For this reason, all comparisons of IID values were made between groups of the same size.

To calculate the measure, the IIDs of all the fish are sorted from least to greatest. If the distribution of IIDs is relatively consistent, implying that all the fish are part of the same group, then the increase in IID between successively ranked fish will be close to constant. Any sudden increase in the series of ranked IIDs implies the presence of a sub-group that is farther from the main group than the members of the main group are from each other. Any fish ranked beyond the IID leap are members of this sub-group. One advantage of this measure is immediately apparent: if two fish split off from the main shoal and form a sub-group, their IIDs will increase and the sub-group will be detected, despite both fish potentially having very small NNDs (as they will be each other’s nearest neighbor). In this manner, our measure takes into account the internal spacing of the shoal. However, the measure still relies on an arbitrary threshold value determined by the experimenter: the maximum permitted size of a jump in the ranked list of IIDs. We set this threshold at the square-root of the mean IID, so that it also varies with the density of the group (i.e. the permitted variance in distances within the shoal depends on the mean distance between fish within the shoal).

3.1.3 Aoki’s measure. All of the measures discussed above rely on some experimenter-determined threshold value for deciding whether or not a fish belongs to a shoal (or which of several shoals it belongs to). It would, of course, be better to define a measure that is, as far as possible, independent of such decisions. Interestingly, the earliest published measure of shoal
membership, as far as I know, performs just such a feat. Yet, despite the dominant influence of Aoki’s (1980) groundbreaking monograph, his measures and analysis techniques have not been adopted by almost anyone.

Aoki (1980) examined the distributions of NND in shoals of field gudgeon of sizes from 2 to 8. He noted that, in trials where the shoal split in two (determined by eye) for a sufficiently large portion of the session, the distribution of NND values became bimodal. Aoki used the depression between the two peaks of the distribution as the threshold for determining the extent of the shoal. In other words, he assumed that an individual’s NNDs are normally (unimodally) distributed when it is within a shoal and a second NND distribution peak therefore represents cases in which the fish is not part of the shoal. It remains unexplained why NNDs of fish that are away from the shoal should nonetheless cluster around a single value – forming a second peak in the distribution – rather than simply giving the distribution a fat, mostly uniform, positive tail. Interestingly, Aoki’s shoal membership threshold of about 44 cm is far larger than 4 BL (Gnathopogon are about the same size as zebrafish, 3-5 cm). In addition, the criterion can only be applied in cases where the shoal splits for a sufficiently large proportion of the session to create the second distribution peak.

3.2 The current measure

All of the criteria described in the literature to date share one assumption: that membership in the shoal is determined ‘intentionally’ by the subjects. In other words, fish adjust their motion relative to other shoal members in such a way as to either remain part of or leave the shoal. This (surprisingly cognitive) approach implies that there is a moment of decision at which the behavior of a fish switches from one mode of movement to the other (even if the change in
motion is gradual). The current measure attempts to identify this shift in behavior as nearly as possible and to use it, rather than a distance threshold, to determine shoal membership.

The criterion is based on the NND for cases in which a single fish leaves the shoal. More complex situations, in which a sub-group of more than one departs the main shoal, are discussed later. We begin by examining the time-series of the NND of a single individual (Figure 3.1 A). As the figure clearly shows, NND fluctuates across time (as also noted by Aoki, 1980). An excursion of a single fish away from the rest of the shoal will appear on the time-series of NND as an extended section of high values. The challenge, therefore, is to determine which values are ‘high’ and which are not.

To begin, the trajectory of each individual is partitioned into movement segments, based on the values of that individual’s NND. To find the beginning and end of each movement segment, we construct the overall distribution of NND values for the session (for all fish). A normal distribution is fit to the NND data and the mode of the best-fit distribution is recorded (Figure 3.1 B). The time-series of each individual’s NND is then partitioned using the mode of the distribution as a threshold (see Figure 3.1 A). Every time the NND of a fish exceeds the mode, a movement segment begins; when the NND falls below the mode, that segment ends.

Each movement segment is now a candidate for being a ‘true’ excursion, an episode during which the fish left the shoal. We assume that only movement segments during which the fish attained a large distance from its nearest neighbor qualify. We must now, again, determine what constitutes a ‘large’ distance and once again this is achieved by turning to a density distribution.

---

4 In this section and elsewhere I refer to the modes of distributions rather than the means since, for many of my measures which are constrained to take only positive values, the mode and the mean do not coincide, even when the measure appears to be normally distributed.
Each movement segment is characterized by the maximal NND attained during that segment, denoted MaxD. The distribution of MaxD for all movement segments is then constructed (Figure 3.1 C). It is assumed that most of the movement segments that comprise this distribution are not excursions but merely represent the constant jostling for position that occurs within a shoal. Only the few movement segments whose MaxD falls outside the main distribution represent excursions. A normal distribution is again fit to the data (the distribution of MaxD). The traditional cutoff of p < 0.05 is used. Thus, only movement segments whose MaxD falls above the 0.05 limit of the distribution are considered excursions.

Two important points need to be made about the procedure for calculating the current shoal-membership criterion. First, like Aoki’s measure, to which it is most similar, it suffers from an unavoidable circularity of definition: deciding whether or not a particular fish is a member of a particular shoal depends on the distribution of distances between all the fish present, including those that will eventually be deemed not members of the shoal. The current measure also faces a related issue: as the threshold for ‘true’ excursions is determined by the exclusion of movement segments in the main distribution of MaxD, the measure will fail if fish spend so long on excursions that the larger part of the distribution reflects excursions rather than non-exursion data. In other words, the measure will only operate effectively where a well-defined shoal exists. When shoaling disappears completely the distribution of MaxD, in a vain effort to detect a shoal where none exists, will stretch itself out and subsume all excursions within the main distribution.

Second, a movement segment whose MaxD is sufficiently large is designated an excursion in toto. Thus, the excursion begins – and the fish is considered no longer part of the shoal – as soon as the movement segment begins, which is whenever the individual’s NND exceeds the mode of the overall distribution of NND (Figure 3.1). Thus, a fish can be on an excursion even when its
NND is smaller than those of some of its shoal-mates who are not on excursions. The beginning of the movement segment is selected as the beginning of the excursion in the belief that this most nearly corresponds to the moment of decision to leave the shoal. Since we expect that the decision to depart is accompanied by a measurable change in behavior, our assumption can be tested directly by comparing the trajectory of the fish to either side of the putative moment of decision. Results of such an analysis are presented below (Section 3.3.5) by way of a partial validation of the current measure.

All of the above relates to a situation in which a single fish departs the main shoal. However, in large shoals it is quite likely that several fish may form their own shoal and leave the original group *en masse*. In such a situation, the NND of each fish would remain deceptively small and the criterion needs to be modified to deal with such cases. Intuitively, a fish that is a member of a splinter-shoal with two members will have a small NND\(_1\) (the distance to the other member of its shoal) and a large NND\(_2\) (the distance to the closest member of the main shoal). The same method employed above can be used to determine a threshold for considering a NND\(_2\) ‘large’.

Thus, for detecting sub-groups of 2 members, we use the same procedure as above but replace NND\(_1\) with a new value: NND\(_2\) – NND\(_1\). The same logic is applied to sub-groups of 3, 4, and on, up to half the number of fish in the tank (if more than half the fish split from the main group they become, by definition, the main group).

To summarize, the current shoal membership criterion is calculated as follows. Define NND\(_0\) of a fish \(= 0\). Then:

1. For each \(n\) from 1 to \(N/2\) (where \(N\) is the number of fish in the tank), construct the overall distribution of \(d\text{NND} = \text{NND}_n - \text{NND}_{n-1}\). Find the mode of this distribution, \(\mu_d\).
2. Partition the time series of dNND for each fish into movement segments, each of which begins when dNND rises above \( \mu_d \) and ends when it falls below \( \mu_d \);

3. Find each segment’s MaxD, the maximal value of dNND attained during that segment.

4. Construct the distribution of MaxD. Find the \( p < 0.05 \) cutoff on the positive side.

5. Any movement segments whose MaxD is larger than the cutoff value are ‘true’ excursions.

This procedure is summarized in Figure 3.1.

3.3 Results

Excursions were calculated for every individual in each dataset. Excursion data for two experiments, described below, are presented in this chapter. Further analyses of the same data are given in the following chapter. Excursion data for the remaining experiments are discussed in Chapter 5.

Experiment 1: Three strains of zebrafish were used: LF, SF, and AB. 8 groups per strain, consisting of 8 fish per group, were each tested daily for 5 consecutive days. Each session lasted for 30 min, of which min 5-10 were coded. All other methods were as described in the previous chapter. Two files could not be tracked due to corrupted video files and were excluded from all analyses. In one AB group, one of the fish died before the last testing day. This group was tested with 7 fish on the last day and all analyses were adjusted accordingly.

Experiment 2: 8 groups, of 8 fish per group, of LF zebrafish were tested once each. Each session lasted for 4 hours, of which min 5-10 of each hour (i.e. min 5-10, 65-70, 125-130, and 185-190) were coded. Thus, the first coding period corresponded exactly (in the experience of the fish) to
the first testing day of Experiment 1. All other methods were as described in the previous chapter.

Both these experiments, as well as all the experiments presented in following chapters, were conducted in the same testing tank, described in detail in Chapter 2. The tank is large, bare, and well-lit and is probably a highly aversive environment when first encountered. With increased exposure to the tank – either across days in Experiment 1 or across hours in Experiment 2 – previous experience with fish (and other animals) leads us to expect them to habituate to the environment. Apart from characterizing the comings and goings in undisturbed zebrafish shoals generally, this chapter and the following one may be considered an exploration of how habituation is expressed in zebrafish shoaling.

Note that, under the current shoal membership criterion, there are several types of excursions that can occur, which vary on the number of fish participating (i.e. the size of the sub-group formed). There are N/2 types of excursions, where N is the number of fish in the tank. For all the experiments discussed in this chapter, N = 8 and there are thus 4 types of excursions. Once the beginnings and ends of the excursions in a particular dataset had been identified, several features of the excursions were examined (sections 3.3.1-3.3.4). Then, the excursion data were used to examine differences in movement characteristics between fish on excursions and fish in the main shoal (section 3.3.5).

3.3.1 Number of excursions. We first examined the simplest excursion-related measure, the number of excursions of each type that occurred during a session. Figure 3.2 shows the mean number of excursions by day for each of the three strains used in Experiment 1 (panel A) and by hour for Experiment 2 (panel B). The first clearly apparent trend is that the mean number of excursions increases across days and, more weakly, across hours of exposure to the testing tank.
The effect of day was confirmed by a repeated measures ANOVA (F(4,76) = 7.84, p < 0.001). None of the other comparisons for either experiment was significant (all p > 0.02) nor was there any effect of strain (p = 0.443).

3.3.2 Duration of excursions. We next examined the distributions of the durations of excursions for the same two experiments. Distributions were compared using a KS test (not modified, in this case, as the duration of each excursion is independent of the durations of other excursions). Note that a significant result on the KS test only means that there is some difference between the distributions, not necessarily that their modes are different. Overall, excursion durations were different between the three strains in Experiment 1 (all p < 0.006), the primary difference being that LF fish tended to perform longer excursions than the other two strains (Figure B.3).

Figure 3.3 shows the distributions of excursion durations for each day of Experiment 1 (panel A; distributions for SF are shown; the AB and LF data followed the same pattern as the SF data) and each hour of Experiment 2 (panel B). In all strains, mean excursion duration increased across days of the experiment and durations became more variable (Table B.1, columns μ and σ; Figure 3.3 A). In addition, excursions in later hours of Experiment 2 were of longer duration than those performed in the first hour of the session (all p < 0.002) and excursions in the third hour were significantly longer than those from hours 2 or 4 (both p < 0.001).

As mentioned above, excursions in the current experiments can be of four types, depending on the number of fish participating in the excursion. Comparisons of excursion duration by excursion type (Table B.2 and Figure B.4) for Experiment 1 revealed that type 1 excursions (a single fish leaving the shoal) tended to be of longer duration than all other excursion types (all p < 0.001).
3.3.3 Summed Group Size. The most commonly reported measure of shoal membership is the distribution of group sizes, usually determined using Pitcher et al.’s (1983) EGS (e.g. Pitcher & Parrish, 1993). Group size is usually defined as the (numerical) size of the group that each fish belongs to. However, by the current shoal-membership criterion a fish may be considered a member of several groups of different sizes at the same time. In other words, it is possible for a fish to be on an excursion of type 1 and an excursion of type 2 at the same time. This is a result of the independence of excursions of each type from each other: type 1 excursions depend only on values of NND₁; type 2 on NND₂-NND₁, and so on. Thus, to create a group size distribution comparable to those reported by other authors, a slightly different scale from theirs was constructed. For each time point, all the excursions of each type taking place at that moment were counted. Excursions of type 1 were given a weight of 1; excursions of type 2 a weight of 0.5; type 3, 0.33; and so on (i.e. each type has a weight of 1/T, where T is the excursion type).

The number of excursions taking place at each moment, each multiplied by its weight coefficient, was summed and the resulting value referred to as the Summed Group Size (SGS). SGS is designed so that it reflects our intuitions concerning the dissolution of a group: when all the fish are in one group, SGS will be 0, as no excursions of any type are occurring. At the opposite end of the scale, if each fish is in its own group, of size 1, SGS will be N (for the experiments reported here, 8). If one fish leaves the main shoal, SGS will be 1 (one excursion of type 1 is taking place). If two fish leave together, SGS also equals 1 (two excursions of type 2) but if each leaves separately, SGS is 2 (two excursions of type 1). Since each fish can be on several types of excursions concurrently, the theoretical maximum of the measure is $SGS_{\text{Max}} = \sum_{i=1}^{N} \frac{N}{i}$ (for $N = 8$, $SGS_{\text{Max}} = 16.64$). It is difficult to imagine, however, a distribution of fish that would yield such a value and, in practice, values above N were rarely observed. Thus, SGS
is constructed such that higher values represent shoals that have broken apart into smaller groups and lower values represent shoals that are more cohesive.

Figure 3.4 present the SGS distributions by day for Experiment 1 (panel A; data for SF are shown; AB and LF data followed the same pattern) and by hour for Experiment 2 (panel B). Detailed results for Experiment 1 are given in Table B.3. There was no significant difference between the strains (all $p > 0.129$). As with excursion durations, SGS increases steadily across days as shoals gradually disperse. SGS also increases between hour 1 and later hours in Experiment 2 (all $p < 0.001$). The SGS distribution for hour 2 was also significantly different from that of hour 4 ($p < 0.001$). No other comparisons were significant.

3.3.4 Correlations between excursion measures. The measures described above all relate to excursions of fish away from the main shoal. As such, it is likely that these measures are capturing only slightly different aspects of the same phenomenon. To examine whether this is the case, cross-correlations were calculated and a principal components analysis (PCA) conducted on all the excursion-related measures. Both analyses were performed using SPSS (v 17.0). The three measures discussed above participated in the analysis: the number of excursions per session (Section 3.3.1); the mean duration of the excursions per session (Section 3.3.2)$^5$; and the mode of the distribution of SGS for each session (Section 3.3.3).

Table 3.1 presents the correlation and PCA results. All but one of the correlations – between number and duration of excursions – were significant, suggesting that there may indeed be a

---

$^5$ The mean duration per session was used for the correlation – rather than the mode of the distribution as above – since a distribution could not be constructed for each session individually, due to the small number of excursions in some sessions. For SGS, which is measured in each frame, a distribution was constructed for each session and the mode of that distribution was used in the analysis.
single process affecting all the measures. The PCA uncovered only one significant component, onto which all the measures were significantly loaded (Table 3.1, Column PCA), and which explained 70.65% of the total variance.

3.3.5 Behavioral changes during excursions. The current shoal membership criterion is predicated on the identification of the beginning of each movement segment that will eventually lead the fish far enough away from the shoal to be considered an excursion. If there is a moment of decision at which a fish begins its departure from the shoal, there may be a corresponding change in the behavior of the fish at that same moment. Any differences observed between an individual’s behavior before and during an excursion would provide both a validation of the shoal-membership criterion and information on the manner in which excursions are initiated.

Comparisons were made of the behavior of individuals immediately before and immediately after they began an excursion. Since the nature of the behavioral shift that signals the decision to leave is unknown, four different behavioral measures were examined: speed, turning rate, polarization, and relative bearing. Distances to shoal mates (NND or IID) were not selected as the shoal membership criterion relies on an increase in NND to identify excursions in the first place. For each measure, one second of data was taken from just before the beginning of an excursion and one second of data following the beginning of the excursion (i.e. during the excursion).

Speed (in cm/sec) was measured as described in Chapter 2. Turning rate was defined as the summed angle (in degrees/sec) through which a fish turned, as derived from the headings of the fish. Polarization, unlike its namesake described in Chapter 2, was measured as the angular
deviation of an individual’s heading from the mean heading of the group. Relative bearing refers to the deviation of the heading of a focal individual from the heading it would have had to take to swim directly into the center of the group. In other words, relative bearing is how far away the fish is from heading towards the centroid of the group, in degrees. These four measures were selected in the belief that they were the most likely to show changes corresponding to the start of an excursion, i.e. that in order to leave the shoal fish would increase or decrease their speed, perform a turn away from the group – and thus likely turn away from the center of the group – and cease to swim in the same direction as their shoal-mates (i.e., decrease their polarization).

Whilst it is possible to construct summed ‘before’ and ‘after’ distributions of each measure and compare them (as above), in this case we have a specific prediction: that the distribution of the measure will change (within subject) from before the excursion to after it begins in some consistent direction. Thus, a more powerful and directed statistical analysis was used.

Distributions of each measure both before and during the excursion were constructed for each dataset individually and the mode of each distribution recorded. Then, the modes of each ‘before’ and ‘after’ distribution were compared using a Wilcoxon signed-rank test (performed using SPSS, v 17.0).

Of the four measures, only the speeds of AB and SF zebrafish were significantly different before and during an excursion (Figure B.5; AB, Z = -4.256, p < 0.001; SF, Z = -3.992, p <0.001; LF, Z = -1.116, p = 0.264). Individual speed was consistently higher during the first second of the excursion than during the second that preceded the excursion. Thus, zebrafish (except for LF)

---

6 The standard measure of polarization, as described in Chapter 2, is a characteristic of groups only, not individuals.
tend to speed up when leaving a shoal. No significant change was found in bearing relative to the
center of the shoal (all \( p > 0.21 \)), polarization (all \( p > 0.12 \)), or turning rate (all \( p > 0.43 \)). It is
possible either that these measures do not change in a consistent direction during an excursion or
that any excursion-related changes in them occurred outside the quite narrow window of time
that was examined here.

Finally, one more facet of the mechanism by which excursions occur was examined: the position
in the shoal from which individuals leave. Distributions were constructed of the ranks (see
Chapter 2, Section 2.3.7) of individual fish just previous to the beginning of an excursion. Figure
3.5 displays the distributions of the ranks of fish about to leave on an excursion, for each strain.
As the overall distribution of ranks is uniform by definition (every rank must be occupied in
every frame), the distributions were compared to a uniform distribution as well as to each other.
All distributions were significantly different from uniform (all \( p < 0.001 \)), implying that fish
preferentially leave the shoal from a particular position (if fish were equally likely to leave from
any rank, the distributions would be uniform). As Figure 3.5 shows, zebrafish display a tendency
to begin excursions from near the back of the shoal (more precisely, the probability of being at
the back of the shoal when an excursion begins is higher than the probability of being in the
center or front), which may result from one of two mechanisms: either fish that spend time at the
back of the shoal are more likely to leave the shoal – possibly for reasons not directly related to
their rank – or leaving the shoal is most often accomplished by first moving to the back of the
shoal. No consistent difference was found across days or hours, or between strains. No other
comparisons were significant.
3.4 Discussion

Most published discussions of individuals leaving groups or of groups splitting occur in the context of so-called fission-fusion societies, most often concern animals in their natural habitats, and involve much larger spatial and temporal scales than in the present case (e.g. Krause & Ruxton, 2002). An analysis of group cohesion at the current scale (both temporal and spatial) has not previously been attempted for any species, as far as I am aware. It is not known whether zebrafish shoals in the wild exchange members more permanently, like other cyprinids (e.g. guppies; Croft et al., 2003) but, if they do, we would not expect to observe such effects in the current experiments. Pays et al. (2007), working with roe deer (*Capreolus capreolus*), used the likelihood of a recently fissioned group to re-fuse as a measure of whether or not the daughter groups had completely separated. Yet the limited space available in the testing tank used here ensures that all excursions end in the return of the fish to the main shoal. In addition, the total number of fish used in these experiments is controlled and small. It is thus possible to argue that the shoals in the experiments described here could not split in any meaningful sense either because there is not enough space for them to do so or because the total number of fish is too small. There is no particular reason, however, to believe that being part of a shoal is a unitary single-level phenomenon. It is quite possible that zebrafish concurrently engage in the types of excursions presented above as well as – space permitting – more permanent fission-fusion of their shoals.

Different excursion-related measures were highly correlated with each other, implying that they mostly reflect different aspects of a single process. It is particularly interesting that, of all the measures, the number of excursions and their duration did not significantly correlate – despite the fact that both values were shown to increase across days – suggesting that zebrafish shoals
break apart on each of these dimensions independently. In other words, a shoal may dissolve because excursions get longer or because there are more of them or (but not necessarily) both.

The experiments presented here involved either repeated (Experiment 1) or long-term (Experiment 2) exposure of the fish to an initially aversive environment. Many of the changes observed across days or hours can be attributed to habituation of the fish to the testing tank. If, as behavioral ecologists have argued, cohesive shoaling is primarily an anti-predatory response (Krause & Ruxton, 2002) – and thus likely to be elicited by the brightly-lit, bare, testing tank – then the progressive dissolution of the shoal across days or hours as presented above is precisely the effect of habituation we would expect to observe as the fish become less fearful of the tank.

The data presented above demonstrate some aspects of the mechanism by which excursions operate. Individual fish departing the shoal alone tend to go on longer excursions than do larger sub-groups (Figure B.4). Excursions most often originate at the back of the shoal (Figure 3.5) and are characterized by an increase in the speed of the departing fish (Figure B.5). The three strains of zebrafish tested behaved, for the most part, similarly to each other. LF zebrafish perform slightly longer excursions than SF or AB fish (Figure B.3) but on all other measures there was no significant difference between the strains.

Thus, zebrafish shoals are dynamic, constantly losing and regaining members on a timescale of just a few seconds to tens of seconds. As the fish habituate to their environment, these excursions increase in both number and duration and may gradually lead to the dissolution of the shoal (Figure 3.4). It is possible that, in a larger tank with more zebrafish, excursions would result in the departing fish leaving the shoal more permanently and joining a different shoal. Such behavior has been observed in guppies on similar timescales (Croft et al., 2003).
The shoal membership criterion developed here is applicable to any group-living species. Several authors have noted that the composition of animal groups is rarely stable, even on the timescales examined here (e.g. Croft et al., 2003) and applying to other species a similar analysis to the one presented above may uncover interesting differences and similarities in the manner in which they order their participation in shoals, flocks, or herds.
Table 3.1. Cross-correlations and results of PCA on excursion-related measures of shoaling.

Correlation tables and significance values of the correlations are presented for all three shoaling measures examined in this chapter. Num E – number of excursions; Dur E – mean duration of excursions; SGS – mode of SGS distribution; r – Pearson cross-correlation coefficient; p – significance of the correlation coefficient; PCA – loadings of each measure onto the one factor extracted by a principal components analysis that included all the measures (see text for details). Non-significant p values are highlighted.

<table>
<thead>
<tr>
<th>Measure</th>
<th>Dur E</th>
<th>SGS</th>
<th>Dur E</th>
<th>SGS</th>
<th>PCA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Num E</td>
<td>0.232</td>
<td>0.827</td>
<td>0.011</td>
<td>&lt;0.001</td>
<td>0.855</td>
</tr>
<tr>
<td>Dur E</td>
<td></td>
<td>0.567</td>
<td></td>
<td>&lt;0.001</td>
<td>0.669</td>
</tr>
<tr>
<td>SGS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.970</td>
</tr>
</tbody>
</table>
Figure Captions

*Figure 3.1.* Procedure for calculating the group membership criterion. First, the time-series of the NND for a single fish is calculated (A). The overall distribution of NND for the session (for all fish) is calculated (B), and the mode of the distribution (red line in A and B) is used as a threshold to partition the time-series of NND into movement segments. Each segment is characterized by the greatest NND attained during the segment, MaxD. The distribution of MaxD is determined (C), and the 0.05 cutoff of this distribution (blue line in A and C) is used as the threshold above which NND must rise for a movement segment to be considered an excursion. See text for more details.

*Figure 3.2.* Mean number of excursions per session by day for each of the three strains in Experiment 1 (A) and by hour for Experiment 2 (B). Error bars indicate ± SEM.

*Figure 3.3.* Density distributions of excursion duration by day for the SF shoals in Experiment 1 (A) and by hour for Experiment 2 (B). Distributions of the other two strains in Experiment 1 were similar to the SF data.

*Figure 3.4.* Density distributions of the summed group size (SGS) by day for the SF shoals in Experiment 1 (A) and by hour for Experiment 2 (B). Distributions of the other two strains in Experiment 1 were similar to the SF data.

*Figure 3.5.* Density distributions of position in the shoal before the beginning of an excursion, by strain for Experiment 1. A rank of 1 is assigned to the fish at the front of the shoal; a rank of 8 to the fish at the back of the shoal. Data are taken from the 1 sec preceding the start of each excursion (see text for details).
Figure 3.1

A

NND (cm)

Time (sec)

B

P

NND (cm)

C

P

MaxD (cm)
Figure 3.2

A

Number of Excursions

Day

LF
AB
SF

B

Number of Excursions

Hour
Figure 3.3

A

B
Figure 3.4

A

B
Figure 3.5
Chapter 4: Characterizing Zebrafish Shoaling

This chapter continues the characterization begun in the previous chapter, presenting many different facets of shoaling in an effort to at least partially delimit the types of structures exhibited by groups of zebrafish. As demonstrated below, distances, speeds, and polarizations of shoals are constantly changing. Individuals move within the shoal (Krause, 1993a) and leave the shoal altogether. Under natural conditions, where multiple shoals occupy overlapping ranges, fidelity to a given shoal is of shockingly short duration, at least in guppies (about 10-20 sec; Croft et al., 2003). As shown by Aoki (1980) and Miller & Gerlai (2008), distances between fish in a shoal oscillate on timescales of several seconds. Thus, shoals are complex. In order to robustly detect any consistent changes in shoaling behavior – resulting, for example, from environmental manipulations – it is necessary to classify shoaling in detail on multiple dimensions. Below, I show that a close examination of distributions of shoal polarization is indispensable to understanding the distinction between shoaling and schooling, that the process of habituation of a shoal to its environment can be quantitatively described, and that the power of a detailed description allows detection of even subtle differences between zebrafish strains.

The data analyzed here are of the same two experiments described in the previous chapter (Section 3.3). Four different measures of shoaling were selected (see Chapter 2): NND, IID, speed, and polarization. The NND, IID, and speed values reported are the mean values (across individuals) for each frame of data. Polarization was measured as the circular standard deviation of the bearings of the fish (Appendix A). Note that this measure increases with decreasing orderliness. Thus, groups whose members are all swimming in exactly the same direction are characterized by a polarization score of 0 and less unanimous groups have higher polarization.
scores\textsuperscript{7}. Polarization distributions were calculated both inclusive and exclusive of fish on excursions. No difference of any importance was found between the two measures and only the data excluding fish on excursions are shown.

\section*{4.1 Distributions of basic measures}

Mean values per frame for NND, IID, speed, and polarization were calculated for each dataset. Figure 4.1 shows the distributions of each measure by day for Experiment 1 (SF data are shown; the data for LF and AB fish followed the same pattern as the SF data, except where noted; detailed data for all strains are given in Table B.4). Figure 4.2 shows the distributions by hour for Experiment 2.

As the figures show, mean NND and IID increase across both days and hours of exposure to the testing tank, probably as a result of habituation to the environment, as discussed in the previous chapter. Both distributions, on later days and later hours, display the beginnings of a second peak at a higher value (this peak is even more visible in the NND distribution of LF fish, Figure B.6), much like that described by Aoki (1980) for field gudgeon shoals. Like Aoki, we assume that this peak reflects excursions of fish away from the shoal. Most days were significantly different from each other and the modes of both the NND and IID distributions increased across days (Table B.4, Column $\mu$). In Experiment 2 (Figure 4.2) most of the change in the distance distributions occurred between the first hour and all later times. For both NND and IID, the distribution for the first hour was significantly different from those of hours 2-4 (all $p < 0.001$).

\textsuperscript{7}Theoretically, the polarization score increases to infinity if bearings are maximally uncorrelated. However, in practice values above 3 – corresponding to a summed vector of size 0.011 – are rarely observed. See Appendix A.
The NND (but not the IID) distribution for hour 4 was also significantly different from hours 2 and 3 (both p < 0.008).

Shoals also slowed down significantly across days (Figure 4.1 C) and hours (Figure 4.2 C) of exposure to the tank (detailed data are presented in Table B.5). As with the distance distributions, the change in behavior in Experiment 2 was concentrated in the interval between hours 1 and 2 (p < 0.001) after which there was no further significant change (all p > 0.07).

Finally, the polarization of shoals also increased (meaning that the shoals became less polarized) across both days (Figure 4.1 D) and hours (Figure 4.2 D). Detailed data are presented in Table B.6. As with the other measures, in Experiment 2 the majority of the change occurred after the first hour (all p < 0.001) and all other hours were not significantly different (all p > 0.69). It is interesting to note the gradual formation across time of a second peak in the polarization distributions at higher values (representing more disorganized shoals). The importance of this finding is discussed in greater detail below.

There were significant differences between the three strains of zebrafish on all four measures in Experiment 1. The NND distributions of the strains all differed from each other (all p < 0.007) as did the IID distributions of the LF and AB fish (p < 0.007). It appeared, however, that the modes of the distributions were similar and that the differences between them depended on their long positive tails. The positive tails of distributions related to distance – like NND or IID – are disproportionately affected by fish on excursions, since a part of each excursion takes place, by definition, far from the shoal. Thus, differences in the pattern of excursions between the strains (already noted in the previous chapter) could be responsible for their different overall NNDs and IIDs. Since the exact beginnings and ends of all excursions are known (Chapter 3), it is possible to recalculate any measure excluding fish that are on excursions from consideration. IID
distributions for all strains were recalculated (Figure 4.3; data for SF fish are shown) and, as can be seen, the positive tail of the distribution was greatly reduced, whilst the location of the distribution peak did not change. Thus, excursions account for most or all of the ‘fat’ positive tail of the IID (and NND, not shown) distributions and, when excursions were excluded, there was no significant difference between the strains (all p < 0.026).

There were also significant differences between the strains’ speeds (Figure 4.4 A). As is immediately apparent, LF zebrafish are significantly slower than the other two strains (both p < 0.001), which are not different from each other (p = 0.16). It is possible that the decreased speed of LF fish is due to the long fins for which they are named, which may negatively affect their swimming capabilities compared to the other two strains (see e.g. Plaut, 2000).

Finally, LF shoals were also significantly less polarized than AB or SF shoals (Figure 4.4 B; both p < 0.001). The latter two strains’ distributions were also significantly, but much less, different from each other (p = 0.003).

Many studies of shoaling have reported another measure assumed to be related to the overall organization of the shoal, the bearing to nearest neighbor (see Chapter 2, section 2.3.6). Distributions of bearings have been reported in a number of species and some interesting regularities have emerged. In flocks of starlings, nearest neighbors are almost never found directly in front or behind and are most often found at bearings of 45° horizontally, and a little above or below a focal individual (Ballerini et al., 2008a). In flocks of surf scoters (melanitta

---

The increased value of the distributions at IID = 0 is an artifact of the way the censored data are calculated: in cases where all the fish are on excursions (as occurs, for instance, when the 8-fish shoal splits into two groups of 4 fish each) the standard measure of IID is undefined and an IID value of 0 is assigned for that frame. As the distribution shows, this is not a rare occurrence.
perspicillata) swimming on the surface of the water, however, nearest neighbors are most often found in front or behind (Lukeman et al., 2010). In fish species, large marine species seem to prefer their nearest neighbors to be alongside them, i.e. at bearings of between 45 and 90° (tuna: Partridge et al., 1983; saithe, cod, and herring: Partridge, 1980) whereas small freshwater species most often have their nearest neighbors directly in front or behind them (minnows: Partridge, 1980; field gudgeon: Aoki, 1980).

Nearest neighbor distributions, excluding fish on excursions, for all three strains in Experiment 1 are shown in Figure 4.5. None of the strains was significantly different from any other (all p > 0.85). The distributions appear to have peaks around 0 degrees, implying that nearest neighbors in zebrafish are most likely to be directly in front of a focal individual, as they are in other freshwater species. However, none of the distributions was significantly different from uniform (all p > 0.57), suggesting that zebrafish nearest neighbors are equally likely to be at any bearing to each other. Analyses of the bearing data divided by testing day and by hour (for Experiment 2) found no significant differences. As bearings to nearest neighbors appeared to be uniformly distributed, this measure was excluded from further analyses.

4.2 Correlations between measures

As with the different measures of excursions discussed in the previous chapter, many of the measures presented above may capture different aspects of the same process and, as a result, correlate with each other. Thus, cross-correlations of the measures discussed above were constructed. Correlations were calculated for each dataset (session) independently and the mean correlation was subjected to a t-test for significance (e.g. Howell, 2001). As expected, the two measures of distance, NND and IID, correlated significantly (mean r = 0.644, p < 0.001). Only one other correlation, that between speed and polarization, was significant (mean r = -0.437, p =
The sign of this correlation is negative for all the strains implying that shoals of zebrafish are either slow and un-polarized or fast and polarized (recall that low polarization scores reflect increased polarization). A similar correlation has previously been noted in the shoaling literature (Parrish et al., 2002; Viscido et al., 2004). All other correlations were non-significant (all p > 0.017).

4.3 Dynamics of shoaling measures

All of the measures discussed above characterize the shoal at a given moment in time but, as demonstrated above, shoals constantly disperse and re-form and are anything but constant across even short timescales. Previous research (with the exception – as always – of Aoki, 1980) has almost entirely ignored the dynamics of shoaling. Here, as a complement to the effects of habituation on timescales of hours or days discussed above, I examine in detail how shoaling evolves over very short timescales.

Dynamics, broadly, can take one or both of two forms: periodic oscillations around a fixed mean; and linear (or curvilinear) trends of the mean. This section presents the results of tests for periodic oscillations over short timescales, within the 5 min coded of each session. Details of the analysis procedure, which has not been used before as far as I am aware, are given in Chapter 2.

Time-series analyses for significant oscillations were performed on the same four measures previously introduced: IID, NND, polarization, and speed. The value of interest was the period of each oscillation (the inverse of the frequency, i.e., the time taken – in seconds – for one full

---

9 There is no reason why measures of shoaling could not also exhibit more complex nonlinear dynamics. However, testing for such dynamics – and validating the results – is notoriously difficult and generally requires more data per session than is collected here (see e.g. Abarbanel et al., 1993, for a possible method to uncover such structures). This would be a fruitful area for future research.
oscillation). The NNDs and speeds of LF shoals were found to oscillate significantly more slowly (i.e. the period of the oscillation was higher) than those of AB shoals (both p < 0.003). All other comparisons (between strains, days, and hours) were not significant. Excluding fish on excursions had no significant effect. More importantly, all the distributions were found to be non-uniform (all p < 0.001) with peaks at periods between 10-40 sec (Figure B.7). Since 5 min of data were collected per session, it should have been possible to detect oscillations of periods up to 150 sec (half the length of the dataset). Nonetheless, the majority of detected oscillations were much shorter (Figure B.7), implying that zebrafish shoals modulate their movement on timescales around 10-40 sec. It is also, of course, possible that there are longer-term oscillations, on the order of tens of minutes, that could not be captured by the present analysis.

Periodic oscillations in some measures of shoaling have been described in the literature. Aoki (1980) performed a very similar analysis to the current one on the mean NNDs of shoals of field gudgeon. Aoki reported mostly faster oscillations (with periods of 2-5 sec and 10-20 sec) than are reported here, but his data consisted of shorter time-series, possibly preventing him from detecting longer oscillations. Viscido et al. (2004) reported fluctuations in the polarization of shoals of giant danios but did not analyze them further. Miller & Gerlai (2008), using a different method than the one employed here, described oscillations – almost indistinguishable from those reported here – in the mean IID of zebrafish shoals of 16 individuals, reinforcing the generality and reliability of the current finding.

---

10 If the oscillations detected had been the result of random fluctuations, the probability of any oscillation being of a given period would have described a uniform distribution. Thus, comparing the data to a uniform distribution tested whether or not the oscillation periods were random.
4.4 Discussion

The first striking result to emerge from the data presented above is that most of the distributions of each measure are more-or-less normally distributed (with the exception of polarization, which appears to be bimodally distributed), and that several measures correlate with each other. Thus, shoals of zebrafish occupy a limited range of the parameter space within which we measure shoaling and, most likely, a more limited range than they are physically capable of. For example, there is no reason why a shoal could not be slow and polarized but the data indicate that this rarely occurs. This suggests quite strongly that zebrafish shoaling is subject to ‘traffic rules’ (Parrish & Hamner, 1997) which limit the types of structures observed in the data.

4.4.1 Effects of day, hour, and strain. Mirroring the excursion results reported in the previous chapter, all four measures reacted in a consistent manner to the habituation of the fish to the testing tank, both across days (Experiment 1) and hours (Experiment 2). NND and IID increased steadily, even when excursions were excluded from consideration. This implies that shoals disperse as they habituate not only via an increased number and duration of excursions but also by gradually increasing the distances between the fish that remain in the shoal. Shoals slow down as they habituate and become less polarized (see below).

It is informative to compare the process of habituation to the tank across 5 repeated exposures of 30 min each (Experiment 1) to one exposure of 4 hours (Experiment 2). The changes in shoaling across days appear gradual and continue for at least 5 days (Figure 4.1) whereas zebrafish left in the tank for several hours show little change in their behavior after the first hour (Figure 4.2). It is possible that the stress of being repeatedly handled (when being placed in and removed from the tank) partially explains this difference.
LF zebrafish shoaled differently from AB or SF fish on several dimensions. LF shoals were slower and less polarized than those of the other strains (Figure 4.4) as well as performing longer excursions (see Chapter 3). Some of these differences, particularly those in speed, may be attributed to physical limitations caused by the elongated fins of the LF fish.

4.4.2 Dynamics. All four measures examined oscillated with periods in the range of 10-40 sec, as previously demonstrated by Miller & Gerlai (2008, for IID) in larger zebrafish shoals and by Aoki (1980, for NND) in field gudgeon shoals. Interestingly, the oscillations detected here did not change across days or hours, and did not vary between zebrafish strains. Thus, it appears that they are a constant feature of the way zebrafish shoal. The simplest explanation of oscillations in any measure is that they reflect a delay in the response of one fish to a change in the behavior of another. For example, if zebrafish take some time to adjust their heading to follow a turning conspecific, the polarization of the shoal may decrease (as the leader turns) and then recover (as the follower(s) match the new heading). However, this explanation is unlikely given the timescales of the most frequent oscillations observed here, between about 10 and 40 sec. Hunter (1969) has shown that mackerel react to a change in direction of one shoal member in between 0.1 and 0.5 sec, a whole order of magnitude faster than the oscillations reported here. It is difficult to imagine a shoal maintaining any sort of cohesion if fish regularly took 30 sec to react to their conspecifics. Miller & Gerlai (2008) suggested that the function of oscillations in distance between zebrafish relates to the conflicting demands of predator avoidance – which is assumed to be enhanced by closeness to conspecifics – and foraging success – which is assumed to be enhanced with increasing distance from conspecifics. Periodic variation of the distances in the shoal may be a more successful compromise than selecting a single, suboptimal, distance. It is also possible that oscillating shoals are more stable and can recover more easily from perturbations (for example by an approaching predator) than crystalline shoals in which distances
and speeds do not vary. If this latter explanation has any merit, it is possible that environmental manipulations – such as detection of a predator – will have an effect on the periods of the oscillations. This question is addressed in the following chapter.

4.4.3 Bimodal polarization distributions. The most interesting finding, in my opinion, to emerge from the data above relates to the apparent bimodality of the polarization distributions (e.g. Figure 4.1 D). Most authors agree that differences in polarization determine (or rather, reflect) whether a group of fish is a shoal – a loose aggregation – or a school – a highly polarized and coordinated group (see Chapter 1). Thus, we might expect to find that these two behavioral modes correspond to two distribution modes in polarization, i.e. that the density distribution of polarization does not vary monotonically across its entire range but that highly polarized groups (schools) and mostly non-polarized groups (shoals) are more common than intermediate structures. No study to date, as far as I am aware, has presented or searched for such data.

Polarization distributions were clearly not normal (all $p < 0.001$) and appeared to be bimodal. This was tested using a modified Maximum Likelihood Estimation (MLE) method, as follows: for each polarization distribution, the best-fit Gaussian mixture model with 1, 2, or 3 elements (i.e. peaks) was determined; the likelihood score for each model was calculated and the best model, as determined by the Bayesian Information Criterion (BIC), was selected (see Appendix A for details). In all cases (distributions by strain, day, or hour), the method confirmed that the data were most likely to be bimodally distributed. This result implies that zebrafish exhibit two types of collective motion: a highly polarized one, which we would be tempted to call schooling, and a less polarized one, shoaling. It is now possible to differentiate putative episodes of shoaling from schooling by dividing the data at the point of inflection between the two distribution peaks, similarly to Aoki’s (1980) NND-based criterion for the splitting of shoals.
(this criterion also suffers from the same limitation as Aoki’s: it can only be applied when the data contain sufficient occurrences of both modes of behavior to form a recognizably bimodal distribution).

The point of inflection between the two modes was identified by fitting a bimodal Gaussian model to each data distribution and solving for the point of intersection of the two elements\textsuperscript{11}. The points of inflection were similar for the AB (0.449) and SF (0.499) strains and slightly higher for the LF strain (0.772). These values correspond to summed movement vectors of magnitude 0.78 for the SF, 0.80 for the AB, and 0.68 for the LF strain (if all the fish are perfectly aligned, the summed vector has a magnitude of 1; if they are maximally unaligned, it has a magnitude of 0). These numbers are easier to grasp if we say, with little loss of rigor, that SF zebrafish shoals become schools when they exceed 78% polarization (and similarly for the other strains). No systematic change in the inflection points of the distributions across days was identified.

Of course, as the shoaling and schooling distributions overlap, ascribing a particular frame of data to either distribution based purely on whether it exceeds the point of inflection or not is associated with an error probability. In addition, as the widths of the distributions are not equal (the shoaling distributions invariably have a larger standard deviation), the probability of misattribution is not the same for both modes of movement. Across all the data from Experiment 1, the ‘false positive’ probability for the schooling distribution (i.e. the proportion of that

\textsuperscript{11} The central point of intersection is selected as the distributions, obviously, intersect again at their positive and negative tails, as they both approach 0. A different method is to select the local minimum of the summed model (e.g. Drai et al., 2000). No difference was found between the two methods.
distribution that lies on the shoaling side of the inflection point) was 0.046; that for the shoaling distribution was 0.482. No difference was found between strains or across testing days.

Apart from identifying the level of polarization at which schooling morphs into shoaling, we can also examine the individual distributions of each movement mode. Figure 4.6 displays the modes (i.e. the locations of the peaks) of the separated polarization distributions for shoaling and schooling by day for each strain. Several interesting phenomena are clear from the graph. First, the modes of the schooling distributions do not increase across days whereas the modes of the shoaling distributions do. Second, there are no differences between the strains’ schooling distributions but their shoaling distributions are different. Thus, it seems that there is a qualitative – and not just quantitative – difference between schooling and shoaling, at least as regards polarization. The typical polarization of a school is constant, not only across days but between strains as well, whereas the polarization of a shoal differs between strains and increases across days.

The bimodal distribution of zebrafish shoal polarizations represents a quantifiable behavioral measure of the difference between shoaling and schooling, the first such data reported to date as far as I am aware. It is particularly interesting that no third peak of polarization appears as the groups habituate to the testing tank and that the second, shoaling, peak becomes more sharply defined across days rather than less so. Both of these results imply that the data cannot be accounted for by appealing to a gradual overall lessening of polarization (i.e. that the putative shoaling element is not simply a ‘fat tail’ of an otherwise unimodal distribution) but that the second peak represents a distinct and separate mode of behavior. Of course, we cannot know how the distribution would continue to evolve if the experiment were continued beyond five days and this is an intriguing area requiring much more work. Note that, as the data exclude fish on
excursions, the overall decrease in polarization across days cannot be merely a result of increased dissolution of the group. Rather, in parallel to the increase in the frequency of excursions across days, zebrafish groups also gradually spend more time shoaling than schooling and their shoals – but not their schools – become progressively looser.

These data provide the first empirical support for an unstated theory of shoaling which, as far as I can tell, all authors subscribe to: that the organization of groups of fish can take the form of highly polarized schools or of loose aggregations called shoals, or can dissolve altogether (which we have excluded from our polarization data by design), and that there is no ‘third mode’ of collective motion in fish. I am not aware that researchers working on aggregations of other species (mammals, birds) adhere to any similar distinction and it is an interesting question, deserving of further research, whether groups of fish are the only ones to display several modes of collective motion or whether such flexibility is present in all animal groups and remains to be uncovered, possibly using the approach introduced here, in less intensely studied systems.
Figure Captions

*Figure 4.1.* Density distributions of NND (A), IID (B), speed (C), and polarization (D) by day for Experiment 1. Data for SF fish are shown; the data for the LF and AB fish followed the same pattern.

*Figure 4.2.* Density distributions of NND (A), IID (B), speed (C), and polarization (D) by hour for Experiment 2.

*Figure 4.3.* Density distributions of IID including all fish (blue) and excluding fish on excursions (red) for Experiment 1. Data for SF fish are shown; the data for the LF and AB fish followed the same pattern.

*Figure 4.4.* Density distributions of speed (A) and polarization (B) by strain for Experiment 1. The distributions for LF zebrafish are significantly different from those of SF and AB fish.

*Figure 4.5.* Density distributions of bearing to nearest neighbor by strain for Experiment 1. Note that bearing distributions are circular and wrap around the x-axis.

*Figure 4.6.* Modes of polarization distributions by day for all strains in Experiment 1, separated into schooling (open) and shoaling (solid) components. See text for details.
Figure 4.1

A

B

C

D

Day 1

Day 2

Day 3

Day 4

Day 5
Figure 4.2

A

B

C

D

P

NND (cm)

IID (cm)

Speed (cm/s)

Polarization
Figure 4.3

![Graph showing distribution of IID (cm) with two curves: one for all fish and another for no excursions. The graph plots P against IID (cm).]
Figure 4.4

A

\[ P \]

\[ 0.2 \]

\[ 0.15 \]

\[ 0.1 \]

\[ 0.05 \]

\[ 0.025 \]

\[ \text{Speed (cm/s)} \]

\[ \text{AB} \]

\[ \text{LF} \]

\[ \text{SF} \]

B

\[ P \]

\[ 0.15 \]

\[ 0.125 \]

\[ 0.1 \]

\[ 0.075 \]

\[ 0.05 \]

\[ 0.025 \]

\[ \text{Polarization} \]
Figure 4.5

![Diagram showing data points and lines representing different conditions (AB, LF, SF) across various bearing degrees.](image-url)
Figure 4.6

Polarization distribution mode ($\mu$)
Chapter 5: Ecological Considerations

Most authors agree that the main adaptive function of shoaling is the reduction of predation risk (Krause & Ruxton, 2002), which operates primarily through predator confusion, the ‘many eyes’ effect, and dilution (see Chapter 1). All these effects – dilution in particular – are affected by the size (i.e. number of members) of the shoal, with larger shoals offering proportionally more protection (Landeau & Terborgh, 1986). Zebrafish (and other species) have been shown to prefer to join the larger of two shoals (Pritchard et al., 2001). Shoaling also has both positive and negative effects on foraging success (Krause & Ruxton, 2002): whilst shoaling may aid in the location of patchy resources (Pitcher et al., 1982) and occasionally permit utilization of otherwise unavailable prey (e.g. that are too big for a single individual to capture or consume), conspecifics foraging in close proximity may compete for prey with, interfere with, or steal from an individual. All these effects may be aggravated by being in a cohesive shoal.

Given all of the above, it is to be expected that fish will alter some aspects of their shoaling when hunger levels, predation risk, or the number of fish in the shoal change. Sogard and Olla (1997), for example, have shown that shoals of food-deprived walleye Pollock have larger mean NNDs than fed shoals, that the presentation of a predator model decreases mean NND for at least 1 hour afterwards, and that these effects interact, with hungrier fish responding less to the threat of a predator. Miller & Gerlai (2007) showed that food-deprived zebrafish shoals have increased IIDs when tested in the presence of dispersed food compared to fish tested in the absence of food. Presenting a predator model (a cardboard hawk silhouette flown over the tank) caused IID to first increase rapidly (in what is called a ‘flash expansion’) and then decrease below pre-predator levels (Miller & Gerlai, 2007). Similar results have been reported for other species (Krause & Ruxton, 2002).
Partridge (1980) showed that mean NND decreases as the size of a shoal of minnows increases from 2 to 6 and Partridge et al. (1980) detected a similar effect in shoals of saithe and cod of between 5 and 25 members. Partridge (1981) demonstrated that shoals of saithe numbering more than 15 fish often divided into sub-groups whose headings diverged from one another. Shoals of less than 11 fish never divided. Partridge (1980) suggested that the motivation to shoal, or to remain in a shoal, increases as the number of fish in the shoal increases, and that this may account for the decreased distances between members of larger shoals (note that this is different from, though perhaps related to, a preference for joining the larger of two shoals, discussed in Chapter 1). Functionally, this increased motivation may result in fish remaining, on average, in larger shoals which, as has been noted above, provide better protection than smaller shoals. Though NND is commonly presented as being of primary importance to fish, it is possible that other considerations indirectly determine its correlation with shoal size. For example, Partridge et al. (1983) suggested that shoals of different sizes assume different shapes and these may constrain NND values to decline in larger shoals.

This chapter describes three experiments that examined the effects of manipulating the number of fish in the shoal, hunger, and predation risk on the characteristics of shoaling in zebrafish. Understanding in detail which aspects of shoaling vary under these manipulations may suggest what mechanisms modulate shoaling to match changing environmental conditions. As in the two preceding chapters, excursions, distributions of basic measures, and the dynamics of the trajectory data were analyzed. In all three experiments, AB fish were used. As Chapters 3 and 4 showed, the behavior of AB shoals is almost indistinguishable from that of SF zebrafish, considered the most similar strain to the ancestral wild-type variety (Plaut, 2000). The AB strain has become the most widely used zebrafish in genetic, developmental, and pharmacological
studies and, increasingly, also in behavioral work. Thus, the AB strain is both sufficiently similar to wild-type populations and apparently of greatest interest to the zebrafish research community.

**5.1 Experiment 3: the effect of N**

In the first experiment, the effect of varying the number of fish in the shoal was examined. Shoals of 5, 10, and 20 zebrafish were tested for three consecutive days each. Following Partridge (1980, 1981), it was expected that larger shoals would exhibit increased cohesion (lower NNDs) compared to smaller shoals when united but might also be more likely to split into smaller sub-groups.

**5.1.1 Methods.** 4 groups each of either 5, 10, or 20 AB zebrafish (denoted N5, N10, and N20, respectively) were tested. Each group was placed in the testing tank for 30 min on three consecutive days. All other methods were as described in Chapter 2 (Section 2.1). One fish in one of the N20 groups died before the last day of testing. This group was therefore tested with 19 fish on the final day and all analyses were adjusted accordingly.

**5.1.2 Results.** Durations of excursions increased across days of the experiment in shoals of all sizes (Table B.7, top), as expected based on the data presented in Chapter 3. Also consistent with previous findings, the speed of the shoals decreased across days in all shoal sizes (Table B.7, bottom). There was no significant increase across days in numbers of excursions in any shoal size (all p > 0.26).

Figure 5.1 shows the effects of the number of fish in the shoal on zebrafish shoaling (see also Table B.7). Shoals of 20 fish exhibited shorter excursions than shoals of 10 or 5 (both p < 0.008; Figure 5.1 A) and had a smaller mean NND (both p < 0.001; Figure 5.1 B). Thus, as in other species, larger shoals of zebrafish are more cohesive and shoal more tightly. Shoals of 5 fish had
the lowest IID (both p < 0.001), obviously, since IID is proportional to shoal size. Shoals of 5 zebrafish were also more polarized than shoals of 10, which, in turn, were more polarized than shoals of 20 (all p < 0.001; Figure 5.1 C). In addition, the modes of both the schooling and shoaling polarization distributions (see Chapter 4) were higher in larger shoals (Figure 5.2). Finally, shoals of 10 fish were faster than shoals of 5 or 20 (both p < 0.001), which did not differ from each other (p = 0.19; Figure 5.1 D).

To examine whether larger groups were more likely to fragment, numbers of excursions for all shoal sizes were compared. Only excursions of type 1 (when a solitary fish leaves the shoal) were examined, thus controlling for the greater number of sub-group combinations available to larger shoals (shoals of 5 can only have excursions of types 1 or 2; shoals of 20 can have 10 types of excursions). In addition, the number of type 1 excursions for each group was divided by the N of that group, to standardize the measure. There was no significant effect of shoal size (F(2,9) = 1.46, p = 0.28) nor an interaction between shoal size and day (F(4,18) = 1.81, p = 0.17). Thus, under the current experimental conditions, larger shoals of zebrafish are no more likely to fragment than smaller shoals.

Distances between the fish showed an interaction effect between experimental day and the size of the shoal (see Table B.7). NNDs were higher on day 3 than on previous days in shoals of 5 and 20 fish (all p < 0.001) but were lower than previous days in shoals of 10 individuals (both p < 0.001). Shoal IIDs followed an identical pattern. It therefore appears that shoals of 5 or 20 zebrafish become slightly less cohesive with repeated exposure to the tank, similarly to the shoals of 8 tested in Chapter 4, whereas distances between members of 10 fish shoals decrease.

---

12 Since IID is the mean of the distances between one fish and all others, it increases with increasing shoal size. Unfortunately, it is not possible to simply standardize the IID values (e.g. by dividing by N).
Finally, NND, IID, speed and polarization all oscillated at similar periods to those reported in the previous chapter. There was no effect of testing day or shoal size on the dynamics (all $p > 0.027$). No other comparisons were significant.

5.1.3 Discussion. As in other species, larger shoals of zebrafish have smaller NNDs. Even the most numerous of the shoals came nowhere near to filling the testing tank, making it unlikely that the decrease in NND with increased shoal size was due to tighter ‘packing’ of the fish into the tank. Larger shoals also had shorter excursions but no more of type 1 excursions than smaller shoals. Thus, individual fish are no more likely to leave a large shoal than a small shoal, implying that shoal size does not correlate with shoal stability in zebrafish.

Large shoals were also less polarized than smaller shoals. The polarization distributions (Table B.8) – as in previous data – were mostly bimodal, with the exception of three datasets (N10, days 2 & 3; N20, day 1) which were better described by tri-modal distributions. The third peak of the distribution always occurred at higher values, representing less polarized shoals. Thus, shoals of 10 or 20 fish sometimes had polarization distributions with a second ‘shoaling’ peak or, alternatively, had a very broad shoaling distribution which is not well described by a normal distribution. Note that, in all cases, a clearly defined schooling peak was still present (Table B.8). Interestingly, the modes of both the schooling (i.e. lower) and the shoaling (higher) components of the polarization distributions (ignoring the third peak, when present) increased with increasing shoal size as did, as a result, the point of inflection between the two modes (Figure 5.2). Thus, larger groups of zebrafish are less polarized than smaller groups both when schooling and when shoaling and their shoaling is more variable in its polarization. It is important to reiterate here that the measure used for polarization, the circular standard deviation of the headings of the fish (see appendix A), is not directly affected by shoal size. In other words, the decreased polarization
of larger shoals is not simply an artifact of the greater number of fish in the shoal. It is probably more difficult perceptually for a fish to align with a larger number of shoal-mates but, as several authors have noted, matching orientations with just a few nearest neighbors can lead to highly polarized groups (e.g. Cavagna et al., 2010) and fish may only consider the orientations of a few neighbors regardless of group size (see e.g. Ballerini et al., 2008a).

Finally, surprisingly, shoals of 10 fish swam faster than shoals of 5 or 20 and, across repeated exposures to the tank, became more cohesive (whilst shoals of 5 or 20 become less cohesive; Table B.7). As shown in Chapter 4 for shoals of 8 fish, speed usually decreases and NND increases across days, as a result of habituation to the environment. These data therefore seem to show that shoals of 8 zebrafish are more similar to shoals of 5 than to shoals of 10 and that intermediate-sized shoals (i.e. of 10 fish) habituate more slowly than small (5 or 8) or large (20) shoals. It is not clear why this should be. Partridge et al. (1983) noted that small shoals of tuna are highly polarized as are very large shoals (over 50 individuals) but intermediate-sized shoals are less organized. The authors suggested that shoals of different sizes may take different shapes which impose a more or less polarized state on the fish and may also affect spacing within the shoal.

In summary, distances between members of larger shoals of zebrafish are smaller, as also found in other species. This decrease in NND is accompanied by a decrease in the mean duration of excursions and no change in their number (relative to N), suggesting that larger shoals are not only tighter but also more cohesive overall. However, this cohesion is achieved without the benefit of increased orderliness. Larger shoals are less polarized than smaller shoals, both when schooling and shoaling. It is possible that staying closer to other members of the shoal assists in maintaining shoal cohesion even when the polarization of the shoal is lower, permitting large
groups to ‘school’ even when their polarization is in the ‘shoaling’ range for smaller shoals (i.e. permits the schooling element of the polarization distribution to spread to encompass higher values; see Figure 5.2). Fish may also modulate the number of neighbors that affect their motion as a function of shoal size (or density). Parrish et al. (2002) varied the number of nearest neighbors that their simulated agents paid attention to and noted that, as this number increased, shoal size increased and polarization decreased, matching the data observed here.

### 5.2 Experiment 4: the effect of food

As noted above, food-deprived zebrafish in the presence of food shoal more loosely than similarly deprived fish that are not given food (Miller & Gerlai, 2007). The current experiment expanded these results to examine the interacting effects of food-deprivation and the presence of food on shoaling. Food-deprived (H) or fed (F) fish were tested either in the presence (+) or absence (-) of scattered food for three sessions. Thus, there were 4 groups in the experiment, denoted H+ (food-deprived, tested with food), H- (food-deprived, tested with no food), F+ (fed, tested with food), and F- (fed, tested with no food). After three trials, all groups were food-deprived and given a probe test in the absence of food.

The presence of food, as previously reported, was expected to decrease shoal cohesion. Hunger (as a result of food-deprivation) might also be expected to lead to dissolution of the shoal, as has been demonstrated in other fish species (see above and Krause & Ruxton, 2002). The probe test at the end of the experiment tested for the possibility that, during the experiment, groups tested in the presence of food (+) learned to anticipate the presentation of food in the testing tank, whether they were hungry during the trials (H) or not (F). These two groups would then be expected to display looser shoaling during the probe test – possibly due to active foraging – than groups that had always been tested in the absence of food (-). The effects of anticipating food might also be
expected to interact with the hunger levels of the fish during the trials. Thus, it is possible that hungry fish (H) would more easily (or quickly) learn about the presence of food in the tank than fed fish (F). In the probe test we therefore predicted that the hungry group that expected food (H+) would display the loosest shoals, followed by the previously fed (but now also hungry) expectant group (F+) and that the most cohesive shoals would be those whose members did not expect food, based on their previous experience of the tank (H- and F-).

5.2.1 Methods. 16 shoals of 8 experimentally naïve AB zebrafish each were tested, 4 groups in each condition, for three sessions each. Fish were either food-deprived for 48 hours before each session (groups denoted H) or fed 30 min before each session (F) and were either tested in the presence (+) of 0.03 g of Tetramin flake food (the same food they were fed in their home tanks) evenly scattered across the surface of the testing tank, or in the absence of any food (-). All groups’ sessions were 48 hours apart to allow for 48 hours of food-deprivation between sessions. After the last (third) day of testing, all groups were food-deprived for 48 hours and then given one additional session in the absence of food.

Groups tested with no food (-) were given additional flake food in their home tanks 30 min after the end of each session so that they were fed the same amount per day as the groups tested with food (+). Fed groups (F) were additionally fed the same amount on days when they were not tested (in between sessions) and thus received twice as much food across the duration of the experiment as the food-deprived (H) groups. The testing tank was emptied and washed clean after each with-food group (+) was tested to ensure that no flakes remained in the tank. All other methods were as described in Chapter 2.
One fish in one of the F- groups died before the third day of testing. This group was therefore tested with 7 fish for the last day of testing and the probe day and all analyses were adjusted accordingly.

5.2.2 Results. Figure 5.3 displays the modes of the different measure distributions across days for all groups (note that the results of the tests presented below, however, are for comparisons between the distributions, not their modes). Figure 5.4 presents the full distributions for all groups on the final, probe, day of the experiment.

In all the measures examined, the difference between the H+ (food-deprived, tested with food) fish and all other groups was particularly evident. Excursion durations – as in the experiments reported in previous chapters – increased across the three days of the experiment in all groups (Figure 5.3 A; D1 < D3, all p < 0.001) except in the H+ group (all p > 0.22). NND and IID, also consistent with previous data, increased across days in all groups (Figure 5.3 C, D; NND, all p < 0.009) except H+ (all p > 0.43). Speed decreased across days in all groups (Figure 5.3 B; all p < 0.002) except H+ (all p > 0.22). Polarization also increased across days (i.e. shoals became less polarized; all p < 0.003) except in group H+ (all p > 0.42). Thus, as shoals habituated to the environment, all measures behaved as expected in all groups except H+.

As above, the polarization distributions were examined for bimodality. All but three of the sessions (H-, day 1; H+ days 1 & 3) were best described by a bimodal distribution. As in the previous experiment, the three sessions that were tri-modal all still had well-defined schooling peaks (at similar polarization levels to bimodal sessions) and had their third peak at higher values than normal, representing more variable and disorganized shoaling. Figure 5.5 shows the modes of the schooling and shoaling components of the overall polarization distributions (summed over days) for each group and the points of inflection between them. Note that the shoaling
distribution mode is lowest in the H+ group (though this difference was not significant; all p > 0.023), suggesting either that shoaling in these fish is more similar to schooling (i.e. more polarized) than the shoaling of the other groups or, alternately, that H+ groups spend less time shoaling than do the other groups.

A general effect of the presence of food in the testing tank was also observed (Figure 5.3 A). Excursion durations for both groups tested with food (F+ and H+) were significantly shorter than those for groups F- and H- (all p < 0.001). Both pairs of groups did not differ within themselves (both p > 0.48). Thus, the presence of scattered food increased shoal cohesion (by this one measure), contrary to our predictions. No consistent significant difference between the H and F groups was found.

At the end of the experiment, all groups were tested under the same conditions: food-deprived and with no food present in the tank. Groups previously tested with food might be expected to behave differently from groups that had not been fed during previous sessions. Excursion durations of the former groups (F+ and H+) were shorter than those of the latter (F- and H-; Figure 5.4 A; all p < 0.001). However, as above, the primary difference was between group H+ and all the other groups. Group H+ had lower NND, IID, and polarization (i.e. was more polarized) than the other groups (Figure 5.4 B, D; all measures, all p <0.001), and was faster than all other groups (Figure 5.4 C; all p < 0.001).

5.2.3 Discussion. The shoaling of zebrafish that were food-deprived and tested in the presence of scattered food (H+) did not change across days in the same way as the other groups, implying either that this group did not habituate to the tank or that the effects of active foraging masked the changes commonly due to habituation. Note that group F+, which was also tested in the presence of food, behaved like the other groups (and like the un-manipulated shoals of
Experiment 1), as did group H-, which was food-deprived. Thus, it was the combination of hunger and the availability of food that caused zebrafish shoals to remain cohesive across repeated exposures to the tank. Both hungry groups with no food (H-) and fed groups in the presence of food (F+) behaved no differently from a control group that was fed and given no food (F-).

Most interestingly, the difference between the H+ group and all others persisted on the probe day, when no group had any food. Since, on the probe day, all groups were food-deprived and tested with no food, the only difference between the H+ group and the others was their past experience of the tank. The H+ and F+ groups might both have learned to expect food to be present but only the H+ group behaved differently from the control groups (those that did not expect food, H- and F-) on the probe test. Thus, only the combination of expectation of food and hunger during ‘training’ led to an observable difference in shoaling, implying that fed fish do not learn as much as hungry fish about the presence of food. Similar results have been reported for other fish species (see Warburton, 2003).

One confusing aspect of these results was the direction of the difference between the groups. Based on the literature reviewed above (see also below), it was expected that hungry shoals in the presence of food (i.e. H+) would disperse – to forage – more than groups that were either not hungry (F+) or had no food to forage on (H-). However, H+ shoals were more cohesive (denser and more polarized) than the other groups. This result may depend partially on the method of presentation of the food flakes, which were broadcast evenly across the surface of the water. This may have discouraged dissolution of the shoal for some previously unsuspected reason, though this result runs counter to the little research that has been done in this field. Several studies have shown that foraging behavior is strongly dependent on the distribution (or patterning) of food,
i.e. whether it is scattered or clumped (e.g. Ryer & Olla, 1991). Zebrafish are known to attempt to defend stable localized food sources from competing conspecifics (Hamilton & Dill, 2002) but nothing is known about their behavior when food is scattered, as in the current experiment. Coho salmon (*Oncorhynchus kisutch*) fry will also defend a localized food source but will not attempt to monopolize access to scattered food (Ryer & Olla, 1996). Ryer & Olla (1995) showed that walleye Pollock foraged individually when food was scattered and in groups when food was clumped and suggested that large clumps of food encouraged cooperation and local enhancement (locating food based on the reactions of conspecifics) whereas scattered food was more likely to lead to competition and thus encourage shoal dissolution. All these data point in the opposite direction to the results presented here and further work is required to delineate exactly under which resource patterns zebrafish shoals disperse or cohere and the contributions of other environmental conditions (such as predation risk; see Ryer & Olla, 1998; Sogard & Olla, 1997).

Finally, it is also possible that the aversive nature of the testing tank discouraged shoal dissolution (e.g. Hamilton & Dill, 2002), though it is not clear why the other groups (and fish in previous experiments, which used an identical method) were not equally affected.

### 5.3 Experiment 5: the effect of a predator

Shoaling is considered by most researchers to have primarily anti-predatory benefits. As such, it is expected that the appearance of a predator stimulus will cause the shoal to become more cohesive and polarized, as this should serve to increase predator confusion (see above and Chapter 1). In addition, it would not be surprising if fish were able to learn where and when predators are likely to appear and might alter their shoaling in anticipation of an attack. Data to support such a suggestion have been collected in related species (e.g. minnows; Magurran & Seghers, 1991). Miller & Gerlai (2007) have shown that zebrafish shoals presented with a
The present experiment was designed as a complement (or contrast) to the previous experiment. During the first three days of the experiment, half the shoals were exposed to a predator image four times during their 30 min session (P) and half were not (N). Then, on the fourth day, half of each group (P or N) were exposed to a predator (+) and half were not (-). Thus, mirroring the previous experiment, during the final, probe, day some groups ‘expected’ a predator and saw one (P+), others expected a predator and did not see one (P-), some groups saw a predator for the first time (N+) and others had not seen and did not now see a predator (N-). It was predicted, based on the data reviewed above, that groups that did not expect a predator (N+) would react to the appearance of the predator most (by increasing shoal cohesion). In addition, the expectation of a predator might affect shoaling behavior on the probe trial whether (P+) or not (P-) one was presented.

5.3.1 Methods. As in the previous experiment, 16 shoals of 8 experimentally naïve AB zebrafish each were tested, four groups in each condition. Each group was tested for three consecutive days either with the predator stimulus (P) or with no predator stimulus (N). On the fourth day of the experiment, half the P and N groups were tested with no predator (-) and half with the predator stimulus (+). Thus, there were 4 groups in the experiment: P+ (predator on all days), P- (predator only on days 1-3), N+ (predator only on day 4), and N- (no predator). All fish were fed ad lib. 30 min before each session.

The testing tank was the same as that used in all previous experiments. However, the fluorescent lighting on both sides of the tank was removed for the present experiment. An LCD projector (Epson, PowerLite 820p) connected to a laptop (Acer AspireOne, ZG5) was mounted by the
ceiling directly beside the video camera such that the image projected by the projector covered the entire testing tank. A Microsoft PowerPoint presentation was displayed to the fish via the projector at all times and this provided the only source of light for the tank (Figure B.2).

The predator stimulus consisted of a silhouette of a hawk, sized so that the wingtips of the image extended across the entire width of the testing tank, which crossed the screen in 3 sec twice, in opposite directions, with a 2 sec interval between crossings. For sessions without a predator stimulus, the fish were presented with a blank white screen for the entire 30 min session. In sessions with a predator stimulus, the predator image was presented over the white screen 4 times during the session, at 7.5, 12.5, 17.5, and 22.5 min from the start of the session. As in previous experiments, min 5-10 of the session were coded. However, in sessions where the predator was presented this period was divided into two segments, before and after the predator appeared (MTT was unable to track the fish during the predator presentation). All other methods were as described in Chapter 2.

5.3.2 Results. Figure 5.6 presents the modes of the measure distributions across days for all groups of the experiment (as above, comparisons reported below are between distributions, not modes). Figure 5.7 displays the full distributions for the probe test.

There was no significant difference in the behavior of the no-predator (N) groups across the first three days of the experiment on any measure (Figure 5.6; all p > 0.012). For the P groups, excursion durations decreased after the first day (Figure 5.6 A; both p < 0.001) but there was no significant difference on any other measure (all p > 0.011). Thus, as opposed to the data in all the previously presented experiments, zebrafish shoals in the present study – whether presented with a predator stimulus or not – showed no evidence of habituation with repeated exposures to the tank. Additionally, the time period before the predator presentation was compared to the period
following the predator’s disappearance in the predator groups (P) for all days but no significant effect was found on any measure (all $p > 0.04$).

On the fourth day of the experiment (the probe day), groups that had previously been exposed to the predator image (P+) had a significantly lower NND during the period before the predator presentation than groups that were later exposed to the predator for the first time ($N+$; $p < 0.001$; Figure 5.7 B). However, no comparable difference was observed involving either the other groups that should have expected a predator (P-) or groups that did not see a predator at all (N-) which had, until that point, identical experiences (i.e., there was no difference between P+ and P- or between N+ and N- until the final predator presentation). The selectivity of the significant difference between groups therefore casts some doubt over the reliability of the result. No other comparisons on any measure were significant.

5.3.3 Discussion. The lack of almost any significant effects in the data presented above can be explained in several ways. It is possible that the slightly altered apparatus (see above) was more aversive than that used for previous experiments. The primary difference was in the lighting which, in the present experiment, came not from fluorescent lights at the sides of the tank but more directly from a projector mounted above the tank. This explanation is supported by the lack of change across days in the behavior of the groups that were not exposed to the predator (N). These groups should have behaved in a similar fashion to control groups in the previous experiments, displaying increased distances and polarization and decreased speeds across the three days of the experiment. That they did not implies that they may have been too fearful of the environment to habituate to it. A similar effect on the groups exposed to a predator image (P) might have masked any effect of the predator. This suggestion is supported by the lack of any
difference between the periods immediately before and immediately after the appearance of the predator, even on the first day of testing.

An alternative, perhaps complementary, explanation is that the predator stimulus was not sufficiently aversive. Miller & Gerlai (2007) – testing shoals of 16 LF zebrafish in the same tank as was used here but under softer lighting conditions – used a cardboard silhouette of a bird flown above the tank to simulate a predator and saw robust anti-predatory responses in their fish. In the current experiment, an animated silhouette was used and the fish exposed to it (P) showed no response, save for a decrease in the duration of their excursions after the first day of the experiment (Figure 5.6 A). This may be due to a lack of response to the predator stimulus, or to a ceiling effect caused by the overly aversive environment, or a combination of both effects.

5.4 Discussion

Shoaling has been proposed to reduce predation via several different mechanisms (see Chapter 1). Some, possibly all, of these mechanisms are sensitive to environmental conditions such as the size of the shoal or the immediacy of a predatory threat. It is therefore interesting to ask to what extent shoaling is responsive to changes in environmental conditions. Here, the methods of characterization developed in the previous chapters were applied to data from three experiments in which the environment of the zebrafish was manipulated.

First, the number of fish in the shoal was changed. As expected from the literature on similar experiments in other species, larger shoals (of 20 fish) were more cohesive though less polarized than smaller shoals (of 10 or 5 fish). Surprisingly, intermediate-sized shoals appeared to be less successful in habituating to the environment than small or large shoals, despite being most similar in number to the shoals used in previous experiments (Chapter 4). Next, hunger and the
presence of food were manipulated, and also found to have an effect on shoaling. Hungry fish that were tested in the presence of scattered food formed more cohesive shoals whose behavior did not change across days like that of fed groups or hungry groups tested without food. Finally, a predator model was presented to the fish. Unfortunately, some unknown aspect of the experimental paradigm (probably related to lighting conditions) affected the fish and masked any differences between the groups. Nonetheless, the difference between the behavior of these fish and those tested in previous experiments revealed that – whatever it was that frightened the fish – shoaling is responsive to fear. Thus, the main result of all three experiments is that the characteristics of shoaling in zebrafish are modified under different environmental conditions. Zebrafish shoals react in measurable ways to the number of conspecifics present, the level of hunger of the fish, the presence or absence of food, and the fearfulness of the fish.

In addition, the direction of these changes in shoaling is consistent and predictable. Fearful responses consist of the elimination of those features associated with habituation in previous experiments; active foraging in food-deprived fish likewise masks the increased laxness of typical habituated shoals.

Finally, Experiments 4 and 5 included a learning element. On the fourth day of each experiment some fish were tested under novel, ‘unexpected’, conditions (either finding no food where some was expected or finding food where none was previously; either seeing a predator for the first time or not seeing an expected predator). In Experiment 4, where all the fish were tested in the absence of food, fish that expected to find food in the tank (H+) remained cohesive, fast, and polarized, as they had during earlier days when food was available, whereas equally deprived fish that did not expect food in the tank (H-) behaved like controls. Most interestingly, food-deprived (for the probe test) fish that expected food to be present but had not been hungry during
previous trials (F+) also behaved like controls, implying that learning about the presence of food depends on both experimental contingencies (the presence of food) and motivation or ‘drive’ (Warburton, 2003).

Similarly, in Experiment 5, fish that expected a predator on the probe day (P+) shoaled more tightly (had a lower NND) even before the predator appeared than fish that did not expect the predator (N+; but see detailed discussion above). Thus, zebrafish not only adapt their shoaling to current environmental conditions but can learn to modulate their behavior to match expected conditions as well.
Figure captions

*Figure 5.1.* The effects of shoal size on zebrafish shoaling. Density distributions of excursion duration (A), NND (B), polarization (C), and speed (D) are shown for shoals of 5 (N5), 10 (N10), and 20 (N20) fish, summed over all three days of Experiment 3.

*Figure 5.2.* Modes of polarization distribution components by shoal size. The modes of the schooling (blue) and shoaling (red) components of the polarization distribution are shown for shoals of 5, 10, and 20 fish. The dotted line shows the point of inflection between the two modes, where schooling becomes shoaling.

*Figure 5.3.* The effects of food and hunger on zebrafish shoaling. Distribution modes of excursion duration (A), speed (B), NND (C), and IID (D) by day for all groups of Experiment 4. F-, fed before each session and tested without food; F+, fed before each session and tested with food; H-, food-deprived before session and tested without food; H+, food-deprived and tested with food.

*Figure 5.4.* Density distributions of excursion duration (A), NND (B), speed (C), and polarization (D) for the final, probe, day for all groups of Experiment 4. Group IDs as in Figure 5.3.

*Figure 5.5.* Modes of polarization distribution components for Experiment 4. The modes of the schooling (blue) and shoaling (red) components of the polarization distribution and the point of inflection between them (yellow) are shown for all groups of Experiment 4, summed over all days. Group IDs as in Figure 5.3.
**Figure 5.6.** The effects of a predator image on zebrafish shoaling. Distribution modes of excursion duration (A), speed (B), NND (C), and IID (D) by day for all groups of Experiment 5. P+, predator image presented on all days of the experiment; P-, predator presented on days 1-3 but not the probe test; N+, predator presented only on the probe test; N-, no predator presented.

**Figure 5.7.** Density distributions of excursion duration (A), NND (B), speed (C), and polarization (D) for the final, probe, day for all groups of Experiment 5. Group IDs as in Figure 5.6.
Figure 5.1
Figure 5.2

The figure shows a graph with three lines representing different categories:
- Blue line with diamonds: School
- Red line with squares: Shoal
- Black dashed line: Inflection

The y-axis is labeled Polarization, ranging from 0 to 1. The x-axis is labeled N, with values 5, 10, and 20.

The graph illustrates the polarization trend across different values of N for each category.
Figure 5.3

A

B

C

D

Excretion duration µ(ser)

$\text{Speed} \, \mu (\text{m/sec})$

NMD µ(cm)

LID µ(cm)

Day

1 2 3 Probe

1 2 3 Probe

1 2 3 Probe

1 2 3 Probe
Figure 5.4
Figure 5.7

A

B

C

D

Excursion duration (sec)

NND (cm)

Speed (cm/sec)

Polarization

P+
P-
N+
N-
Chapter 6: Testing Models of Shoaling

Shoaling research, throughout its history, has attracted a lot of interest from modelers. Breder (1954), following earlier writers, realized that the basic structure of shoaling could be explained by assuming that two ‘social forces’ acted on each agent (simulated fish): an attractive force that pulled the agents towards each other and a repulsive force that prevented them from approaching too close to each other. In most models, these forces are modulated by the distance between the actors and these are commonly referred to as tendency-distance models (Warburton & Lazarus, 1991). Breder (1954)\textsuperscript{13} found that even a simple model incorporating only these two forces, with suitably selected parameters, generated spatial distributions remarkably similar to those observed in a number of different fish species. Many models add a third social force that directs agents to orient in the same direction as their neighbors, called an alignment force. The resulting three-part function is called an Attraction-Alignment-Repulsion (AAR) function. In some models the AAR function is continuous (e.g. Viscido et al., 2004) and in others each force holds sway in its own zone around each agent (e.g. Couzin et al., 2002).

Most models fall into two general categories (Mirabet et al., 2007): Statistical and heuristic (or Eulerian and Lagrangian). Statistical models mostly deal with large-scale behaviors of very large groups, and are not discussed here. An extensive introduction to statistical models and their relationship to heuristic models may be found in Flierl et al. (1999).

The mechanism common to all (heuristic) models assumes that each agent behaves independently according to a set of usually fixed rules that determine its interactions with fellow

\textsuperscript{13} In Breder’s (1954) model, only repulsion depends on distance. The full model is: Cohesion = Attraction – Repulsion/Distance\textsuperscript{2}. 
agents. The simplest models required to generate some form of recognizable shoaling behavior assume that agents are repulsed from other agents that are too close and attracted to agents that are far away. This leads to groups of agents, shoals, that stay together but each member of which also maintains its ‘personal space’.

It is interesting, and relevant to the current discussion, to note that the level on which the vast majority of models attempt to explain shoaling is almost physiological. Few published models have sought to address the effects on shoaling of the factors examined in earlier chapters of this work: habituation, food-deprivation, or the presence of a predator (but see e.g. James et al., 2004; Parrish & Edelstein-Keshet, 1999). This is despite the contention of behavioral ecologists that the adaptive functions of shoaling are precisely these – avoiding predation and enhancing foraging – and such environmental manipulations are thus likely to be the most important determinants of shoaling behavior characteristics (Krause & Ruxton, 2002).

This chapter presents two recent leading tendency-distance models, one zone-based (Couzin et al., 2002) and one continuous (Viscido et al., 2004), and compares them in a novel manner to some of the empirical data presented in Chapters 3 and 4 (i.e. to Experiment 1). In addition, where the models fail most spectacularly, a modest suggestion for improvement is made. A sample model is generated which, while by no means a fully-fledged competitor to existing models, nonetheless captures some behaviors exhibited by zebrafish shoals which they do not (whilst also, of course, failing to capture a lot of the behavior that they do). Comparing models to empirical data, which – due to the scarcity of detailed trajectory data – has not often been accomplished, is sure to both improve the models and change the way we view the data.
6.1 Methods for testing models against data

Both models of shoaling to be examined consist of a set of rules followed by each individual agent. At each simulated time-point, each agent evaluates its ‘perceptual’ input, consisting of the positions and bearings of all the other agents it can detect, and then decides in which direction to move. Thus, the models are essentially heuristic prescriptions for how to behave when in a shoal. To compare the models to the empirical data, the relative positions, speeds, and bearings of each fish were recorded from each frame of data (see Chapter 4, where these data are analyzed). Given the positions of all its shoal-mates and their headings and speeds, the models provide predictions of which direction each fish should move in, if it were following the rules of that particular model (importantly: for a given set of parameter values). Since the actual movement performed by the fish is known, the angular deviation between the prediction and the data can be computed. Finally, a distribution of the deviations of model from data – of errors – is constructed. The narrower and closer to 0° the peak of this distribution is, the more precisely the model describes the data (if the model were perfect, the ideal error distribution would approach the Dirac delta function; if it were useless, a uniform distribution would result).

Theoretically, it should be possible to extend this method to determine the best parameter values to use for each model. The error distribution for any set of parameter values could be compared to the ideal distribution, a sharp peak centered at 0°. The parameters of the model could thus be regressed onto the data, using the error distribution as a measure of goodness-of-fit (or rather, how closely the error distribution approaches the ideal distribution), to find the parameters that best approximate the empirical data. This method was attempted and it was found that the regression invariably returned obviously sub-optimal results. A laborious manual exploration of the models’ parameter spaces revealed that the primary reason for this failure was that the space
itself was ‘bumpy’. In other words, there are any number of parameter combinations that fit the
data better than any of the adjacent combinations, i.e. they inhabit surprisingly steep local
minima of the regression’s parameter space. The regression is ‘trapped’ by one or another of
these local minima, depending on the initial parameter values it is fed, even when advanced
techniques are employed to overcome exactly such a situation (e.g. I used the popular
Levenberg-Marquardt algorithm). Thus, this method was abandoned. Parameter values for the
models were estimated from the publications in which the models were presented and by manual
exploration of the parameter space (see below). It is therefore possible, though somewhat
unlikely, that the models tested here could be made to fit the data more closely than is presented
here by identifying better parameter values than either the authors of the models or I were able to
locate.

6.2 Testing existing models

Two models were chosen to compare to our empirical data, those presented by Couzin et al.
(2002) and Viscido et al. (2004). These two models were selected for several reasons. First, they
represent the two basic types of AAR functions that exist: zone-based and continuous. Second,
these two are amongst the most successful (at generating recognizable shoaling) and most
popular models and have been widely cited. Couzin et al.’s (2002) model, in particular, has
become the basis for a number of elaborations examining more specific questions such as the
role of leadership (Conradt et al., 2009) or differences in directional preference within a shoal
(Couzin et al., 2005). Additionally, the parameter spaces of these two models have been explored
in greater detail than most other models, making it likely that the limits of their possible
behaviors have been fully mapped. Finally, Viscido et al.’s (2004) model has the advantage, for
our purposes, that it was initially designed with shoals of giant danios – a close relative of zebrafish – in mind and calibrated to match empirical data on shoals of 4 or 8 giant danios.

6.2.1 Couzin et al. (2002). Couzin et al.’s (2002) model assumes that each agent is surrounded by three concentric zones (Figure 6.1). Immediately adjacent to each agent is its zone of repulsion (zor), its personal space. If any other agents are detected within zor, the focal agent turns away from them in an attempt to keep its zor empty. This is the cardinal movement rule and outranks both other rules. Beyond zor is the zone of orientation (zoo) and beyond that the zone of attraction (zoa). If zor is empty the focal agent will attempt to match orientations with all detected agents in zoo and (concurrently) to swim towards any detected agents in zoa.

Several limitations also affect the motion of agents in the model: agents are unable to ‘see’ directly behind them, in other words, they have a limited field of view of $\alpha^\circ$; agents can only turn through a certain maximal angle ($\theta^\circ$) in one time-step; and all agents move at the same speed, $s$, at all times. In addition, noise is added to the motion of each agent at each step. The noise is randomly sampled from a Gaussian distribution centered at $0^\circ$ with standard deviation $\sigma^\circ$.

As can be seen, the model has quite a few tunable parameters. Couzin et al. (2002) explored much of the parameter space (see their Table 1): $\alpha$, the angle of vision, was varied between $200^\circ$ and $360^\circ$; $\theta$, the maximal turning rate, between $10^\circ$ and $100^\circ$ per time-step; and $s$, the agents’ speed, between 1 and 5 units per time-step. More interestingly, the radii of the various zones (zor, zoo, and zoa) were also varied, as was the number of agents (from 10 to 100).

6.2.2 Viscido et al. (2004). Viscido et al.’s (2004) model shares several features with Couzin et al.’s (2002) model: here, too, agents can only perceive fish within a fixed viewing angle, set at $150^\circ$; space is assumed to be infinite and empty; and a random vector is added to the one
determined by the movement rules. Most importantly, here, too, agents are subject to three forces: attraction, alignment, and repulsion. In one key aspect, though, Viscido et al.’s model is different from Couzin et al.’s: their agents are subject to just one social force at a time, determined by a continuous AAR function (Figure 6.2). The AAR function determines the force that acts on an agent. Viscido et al. determined a maximal force that an agent could be exposed to ($F_{\text{max}}$), and set it at 12 BL/sec$^2$, a value they obtained from empirical data on shoals of giant danios. The social force was summed with the agent’s previous vector to give the new vector for each agent at each time-point. Viscido et al., unlike Couzin et al., allowed their agents to accelerate and decelerate, but set a maximal speed of 12 BL/sec.

As Figure 6.2 shows, Viscido et al.’s AAR function is actually two functions. When the neighbors of an agent are in the alignment zone, between $\delta_r$ and $\delta_a$ BL away, the function returns neither attraction nor repulsion. Instead, a separate alignment function operates on the agents, identical to the one used by Couzin et al. (2002). The strength of this alignment function was varied between 0.5 and 50 % of $F_{\text{max}}$. The center of the alignment zone was fixed, based on measurements of giant danios, at 1.9 BL. The zone extended 0.5 BL in either direction, i.e. $\delta_r$, the beginning of the repulsion zone, was fixed at 1.4 BL, and $\delta_a$, the beginning of the attraction zone, at 2.4 BL. In addition, the attraction function was designed so that it reached its maximal value ($F_{\text{max}}$), at 5 BL. Agents could only detect each other up to 100 BL away, after which they ceased to influence each other’s movement.

6.2.3 Similarities and differences between the models. The two models described above differ on three key issues. In Couzin et al.’s (2002) model, the influence of an agent on its neighbor depends on which zone that neighbor is in and is entirely independent of where in that zone it is (unless it is outside the agent’s viewing angle). In Viscido et al.’s (2004) model, agents are more...
or less attractive or repulsive to other agents as a function of the distance between them. Only for alignment do both models use the same function, though Viscido et al. modulate the strength of theirs such that it is weaker than the attraction-repulsion force.

The second major difference between the models relates to the hierarchy of their rules. In Viscido et al.’s model, all detected agents affect the motion of a focal agent at all times. In other words, the force acting on an agent is, at all times, the sum of the social forces applied by all other agents closer than 100 BL (suitably scaled so as not to exceed $F_{\text{max}}$). In Couzin et al.’s model, on the other hand, repulsion overcomes the other rules and agents that are swimming away from intruders in their zor ignore any agents in the other two zones.

A final important difference relates to the level of physical verisimilitude of the two models. Viscido et al.’s (2004) rules determine a social force that acts on the agents, which are given a mass (1.7 g) and size ($5.3 \times 0.8$ cm). In a later paper (Viscido et al., 2007) the authors even examined the effects of drag (the resistance of the water) on the results of the same model. Couzin et al.’s (2002) model, on the other hand, does not go as far in simulating the physical environment of the agents (who have, for example, no dimensionality themselves).

6.2.4 Model modifications. In order to model our empirical data, the theoretical models had to be fit to the physical space of the experiments, for instance, where the maximal perception distance is concerned. As the zebrafish were tested in a rather small enclosure (91 cm in diameter), it was assumed that all fish could perceive all other fish at all times, unless they were behind each other, outside the fish’s viewing angle. Certainly, our arena is far less than the 100 BL maximal perception distance assumed by Viscido et al. (2004) and is even slightly less than the 31 BL maximal radius of zoa tested by Couzin et al. (2002).
Agents in both models were free to move within an infinite empty space which, unfortunately, was not the case in our experiments. To avoid differences between the models’ predictions and the empirical data resulting purely from the presence of the walls of the tank, any fish that were less than 10 cm from a wall of the testing tank were excluded from the comparison. Of course, it is possible that the walls of the tank limit the movement of zebrafish from further away than 10 cm. In addition, whilst both models generate 3-dimensional data, 2-dimensional versions were employed, to match the experimental data. The random movement element, the added noise in both models, was discarded. Since the aim of the current comparison is to determine how closely the rules of each model fit the empirical data adding noise to the rules, though it may lead to better-looking behavior of the simulations, was deemed counter-productive. The size of the maximal viewing angle was fixed at 320°, since zebrafish are known to have large visual fields. The preferred NND, set by Viscido et al. (2004) at 1.9 BL, was determined by referring to the data distributions (e.g. Figure 4.1 A). The current comparison only examined the angular precision of the models’ predictions and agent speeds were therefore ignored.

As each dataset from Experiment 1 contains a large amount of data, a small subset of files was chosen to be compared to the models. The 5 datasets from all testing days of one group of SF zebrafish were chosen, as they were deemed the most representative shoal in the experiment. Analyses of the fit between model and data were made separately for each file, permitting a comparison between the fit of the model for different testing days.

Finally, since fish on excursions are likely to be following different rules than fish in the main shoal, all excursions were excluded from the current analysis.

6.2.5 Couzin et al. (2002) results. Initially, before detailed comparisons were attempted, the parameter space of the model was examined in the region considered most likely to provide the
best fit to the data, as determined from the descriptions of the model by its authors. First, a
calibration value needed be selected. Couzin et al.’s model operates in dimensionless units of
distance and speed, whereas our empirical data are in cm. Each distance unit in the model was
therefore considered equal to 1 BL (3.5 cm) in the empirical data.

The only parameters of the original model not specified above, which are of primary interest to
the current analysis, are the radii of the various zones around each agent. For our purposes, \( z_{oa} \)
extends to infinity (or rather, to the edges of the tank), so only the radii of \( z_{or} \) and \( z_{oo} \) remain to
be determined. Couzin et al. fixed the radius of \( z_{or} \) at 1 unit and so the value for our comparisons
was set at 3.5 cm and not varied (informal testing of the effects of varying this value implied that
this was close to the ideal value). To determine the best value for the radius of \( z_{oo} \), which the
authors of the model varied from 0 to 15 units (0 to 52.5 cm), data from the first day of testing
were compared to the model under several values spanning this range.

The error distributions (the distribution of deviations between the model and the data) are
presented in Figure 6.3. As the figure shows, having no \( z_{oo} \) at all (a radius of 0 cm) resulted in
an almost uniform distribution – meaning that the data were equally likely to deviate from the
model by any angle – implying that the model fit the data poorly. As the radius of \( z_{oo} \) was
increased the error distribution became sharper and narrower, improving the fit of the model to
the data. Very little improvement was observed beyond values of about 20 cm. Thus, for all
further comparisons, the radius of \( z_{oo} \) was set at 25 cm.

Figure 6.4 presents the error distributions by day. Clearly, the model – with the current parameter
values – predicted the behavior of the fish on day 1 much better than it did on later days. Indeed,
the error distributions for the fourth and fifth days were not significantly different from uniform
(both \( p > 0.195 \)). Two possible reasons for the degraded fit were considered: either the behavior
of fish that had been exposed to the tank for longer followed a different set of rules (i.e. different parameter values) than did un-habituated fish or the behavior of habituated fish was simply less predictable than that of the fish on the first day. To test the former suggestion, the parameter space of the model was once more explored, this time in comparison to the data from testing day 5. Both the radius of \( z_{oo} \) and of \( z_{or} \) were varied, one at a time, over the same range of values as above (0 to 10.5 cm [3 BL] for \( z_{or} \); 0 to 50 cm for \( z_{oo} \)) but no values were found that significantly improved the fit of model to data.

These results imply that the behavior of the zebrafish on day 5 was not simply under the control of a different set of rules than their behavior on earlier days but that it was inherently less predictable (i.e. more variable). Note that this cannot be as a result of the increase in excursions across testing days as all fish on excursions were excluded from the analysis.

One additional analysis was conducted on the data. Several authors have discussed the differences between fish in different positions in a shoal (e.g. Krause, 1993a) and some have implied that fish in different positions in the shoal may behave differently (e.g. Conradt et al., 2009). Thus, it is possible that the model will fit better or worse when compared to data from individuals in different positions within the shoal. As before (see Chapter 3), the front-to-back rankings of fish within the shoal were used. However, as we wished to exclude fish that were on excursions, the ranking system described in Chapter 2 required some adjustment.

If fish that are on excursions are excluded from the data, the number of fish ranked in each frame will vary. Although the leader of the shoal will always be ranked 1, the last fish in the shoal will receive a different ranking depending on the size of the main shoal at that moment. The ranking measure therefore needs to be standardized so that, for example, being in the center of a shoal is
distinct from being the last fish. Thus, we define \( r = \frac{\text{Rank} - 1}{N - 1} \), where \( N \) is the size of the shoal at that moment. \( r \) ranges from 0 for the leader of the shoal to 1 for the straggler. The central fish (for odd-numbered shoals) always has an \( r \) of 0.5.

Error distributions were constructed (using the data from day 1) separately for different values of \( r \) (Figure 6.5). The range of possible rankings, excluding the leader and last fish in the shoal, was divided into four segments and separate distributions were made for the leader (\( r = 0 \)) and last fish (\( r = 1 \)) in the shoal. As the figure shows, the model predicted the behavior of the leader of the shoal quite well, did not explain the behavior of central fish in the shoal very well, and explained the behavior of fish close to or at the back of the shoal best. This curious pattern of results may derive from two different causes. Generally, as the figure shows, fish that were further back in the shoal behaved more like the model’s predictions than fish that were further forwards. The exception to this rule was the leader of the shoal (for whom \( r = 0 \)) whose behavior was also predicted well by the model. Several authors have noted the difference between the leader of a group and the other members (e.g. recently, Nagy et al., 2010; see also Harcourt et al., 2009). Possibly, in this case, the leader of the shoal followed different rules from the remaining fish – for instance having a larger zor – causing it to swim away from fish that were outside the normal repulsion zone. This may explain why the model predicted the behavior of the leader quite well but not that of the fish directly behind the leader.

6.2.6 Viscido et al. (2004) results. The same comparisons as above were performed on Viscido et al.’s (2004) model. All values were rescaled to be in units of body lengths (BL), as used by the authors. As above, 1 BL was set at 3.5 cm.
To begin, the parameter space of the model was examined. Three parameters were varied (see Figure 6.2), consecutively: $\delta_p$, the preferred NND of each agent; $\delta_m$, the distance at which the attraction function reached its maximal value (denoted $F_{\text{max}}$, which was not varied); and $A_{ij}$, the magnitude of the alignment function. Viscido et al. modulated only the last of these. The authors fixed $\delta_p$, the midpoint of the alignment zone, at 1.9 BL. Since our data imply that zebrafish prefer to remain somewhat closer to each other than that, $\delta_p$ was varied from 1 to 2.5 BL. The width of the alignment zone was left at 1 BL, as in Viscido et al. (2004). The authors’ value for $\delta_p$ of about 2 BL was found to be close to optimal but no better than 1.5 BL, which was closer to the mode of the empirical distribution (3.04 cm; see Figure 4.1 A). Thus, $\delta_p$ was set at 1.5 BL for all remaining comparisons.

Next, $\delta_m$, the distance at which attraction reaches its maximal value, was explored. Note that varying this value, whilst keeping $\delta_p$ and $F_{\text{max}}$ constant, has the incidental effect of changing the slope of the attraction function (Figure 6.2). Viscido et al. set $\delta_m$ at 5 BL (which was also, probably not coincidentally, their threshold for group membership) and did not vary it. However, as the error distributions, presented in Figure 6.6, show, larger values of $\delta_m$ matched the behavior of zebrafish much better than the author’s chosen value. A value of 9 BL was set for all further comparisons.

Finally, $A_{ij}$, the strength of the alignment force was also varied. Viscido et al. varied $A_{ij}$ from 0.5 to 50 % of $F_{\text{max}}$ (which was fixed at 12 BL/sec$^2$). Here, values of $A_{ij}$ from 0 to 100 % of $F_{\text{max}}$ were examined. The fit of the model to the data improved as $A_{ij}$ was increased from 0 to about 60% of $F_{\text{max}}$ and remained more-or-less constant thereafter. $A_{ij}$ was therefore set at 66 % of $F_{\text{max}}$.

Next, the fit of the model to data from different testing days was examined (Figure 6.7). As with Couzin et al.’s model, this model – under the parameters determined above – fit the first two
days well but was not statistically better than random on days 4 and 5 (both \( p > 0.031 \); the
distribution for day 3, despite its appearance, was significantly different from uniform, \( p <
0.001 \)). As above, an attempt was made to improve the fit of the model to the data from day 5 by
searching for better parameter values. All three parameters (\( \delta_p, \delta_m, \) and \( A_{ij} \)) were varied across
the same range of values as above and, as with Couzin et al.’s model, none of the manipulations
improved the fit of the model to the data, lending support to the conclusion that the behavior of
the fish on day 5 was not simply different from their behavior on day 1 but inherently less
predictable.

Finally, the fit of the model to the data was examined by rank. Figure 6.8 displays the error
distributions by rank and shows that Viscido et al.’s model, unlike Couzin et al.’s, predicted the
behavior of fish at all positions in the shoal equally well (all \( p > 0.999 \)).

### 6.3 A modest proposal

Both of the models tested were quite successful at predicting at least some of the empirical data
on zebrafish shoaling. Obviously, as the widths of the error distributions show, there is room for
improvement but it is not possible, unfortunately, to use the error distributions directly to make
targeted suggestions that will lead to such improvement. Both models were much worse at
predicting the empirical data from day 5 of the experiment than they were at predicting the data
from day 1. I have suggested above that this may be due to increasing variability of the behavior
of the fish as they habituate to their environment and thus might be effectively simulated by
increasing the contribution of the random element in both models (which was removed in its
entirety from the current comparisons). However, increasing the randomness of the agents’
behavior in the models might lend the simulations more similar characteristics (such as NND or
polarization) to the empirical data, but would not – except accidentally – improve the fit of the
models to the data under the type of comparison performed above. In other words, adding random noise to a uniform distribution will not make it any less uniform. In addition, this suggestion presupposes that the change in the behavior of the fish across days was truly random, rather than structured in some way that modifications of the models’ parameters simply cannot capture. Perhaps a completely different model structure than the two tested here would be less susceptible to the effects of day and would predict with equal precision the behavior of habituated and un-habituated fish.

The most glaring failure of the models tested, however, is not apparent in the error distributions presented above. Neither of the models – nor any other model I am aware of – generates what I have called excursions (Chapter 3): agents leaving and then rejoining the group (unless the model parameters are set such that the group dissolves, in which case it does so completely and irrevocably). Thus, the most drastic improvement in the similarity of simulated to real shoaling could be achieved, I believe, by allowing individual agents or small sub-groups to leave and then rejoin the main shoal.

Existing models, such as the two presented above, could be adapted with relative ease to permit agents to leave and rejoin the shoal. In Couzin et al.’s model, for instance, this could be achieved by temporarily increasing the radius of *zor* for one or more agents, causing them to be repulsed by the group and to move away from it. This sudden bout of antisociality could occur with a fixed probability at random times. Alternatively, the detailed data presented in previous chapters of this work could be used to calibrate the excursions more finely. For instance, Chapter 3 demonstrates that zebrafish most often leave a shoal from the back. Thus, the probability of leaving the shoal in the models could be made to depend on rank. In addition, the frequency and
duration of excursions increase across days of habituation to the testing environment and this could also be simulated by modulating the probability of leaving.

To examine how excursions might be simulated, a simple model was constructed that focused on leaving and returning to the shoal. In addition to the existence of excursions, this analysis was motivated by the casual observation that fish from the main shoal often swam a little way out of the shoal towards fish on excursions, both when the latter were beginning and ending their excursions (i.e. on their way out of or back into the shoal).

Agents in the current model follow just one of two behavioral rules. First, at every time-point, each agent selects another agent to follow and turns to swim towards that agent. Alternatively, at any time, with probability $P_E$ ($E$ denoting ‘excursion’), an agent may decide not to follow any other agent, in which case it turns to swim away from the rest of the group (using the same rule as Couzin et al.’s zor). Finally, at any time, with probability $P_C$ ($C$ denoting ‘change’), an agent will choose a new agent to follow. Agents that are not following any other agent (i.e. that are leaving the group) are twice as attractive as other agents, in other words are two times as likely to be followed.

These are the only movement rules. The model incorporates no personal space and no AAR function. This is not intended as a criticism of the methods used by almost all existing models; on the contrary, the incorporation of tendency-distance rules into the current model would no doubt improve its correlation to empirical data. However, for simplicity, the single behavioral rule “either follow another agent or go on an excursion” was examined on its own.

To facilitate comparison with the empirical data presented in previous chapters, the parameters of the model were made as similar as possible to the experimental paradigm. Groups of 8 agents
were simulated in each run of the model. Rather than permitting the agents to move in infinite space (as the models above do), the simulations were run in a virtual tank that was circular and of a diameter of 100 cm. Agents did not actively avoid the wall of the tank but, if their vectors intersected the wall, were reflected – along with their momentum – off it, back into the tank. All agents moved with the same speed, set at 4 cm/s to match the empirical data (Figure 4.1 C). Other parameters of the model were set as closely as possible to those used by the other models introduced above. Agents had a visual field of 320° and could turn through a maximum angle of 100°/sec (see Krause & Tegeder, 1994). A random angle – selected from a normal distribution centered at 0 cm and with a standard deviation of 2 cm – was added to each agent’s motion, as in the other models. 10 replications were run for 1000 time-steps each. The values of $P_E$ (probability of going on an excursion) and $P_C$ (probability of changing agent to follow) were set at 0.1 and 0.2, respectively.

The trajectories from the current model cannot be compared to the empirical data in the same way as the two published models were, as the rules followed by agents in the current model change with time. For instance, in order to determine a predicted vector for a fish it would be necessary to know which of its shoal-mates the focal fish was ‘following’. Instead, the simulated data were analyzed in exactly the same way as the empirical data.

Figure 6.9 presents the various measure distributions for the simulated data (see Chapter 4 for definitions and corresponding empirical distributions). Distributions for speed are not presented as all agents were constrained to move at the same fixed speed. As is apparent from the figure, simulated shoals do not cohere very well in the current model. Both the NND and IID values are much higher than they are for the corresponding empirical data (Figure 4.1 A, B). An analysis of excursions performed by the simulated agents (as described in Chapter 3) registered an average
of 128 (± 30.6) excursions of all types per ‘session’, of mean duration 5.4 (± 3.89) sec (compare empirical values from Figures 3.2 and 3.3, respectively). Thus, the current model suffers from the opposite problem to that of most other models, whose agents never leave the closely-knit group. In the current model, agents leave the group too readily and the cohesion of the group is lost\textsuperscript{14}. This lends support to the idea that incorporating the current rule into existing models may result in a good fit with empirical data.

Figure 6.9 C and D show the distributions of polarization for the model data both inclusive of and excluding agents on excursions. Note that the start and end points of excursions were deduced from the trajectory data – as if they were empirical data – and not by observing the rules by which each agent was operating at a given moment. Unsurprisingly, as the model contains no alignment rule, the polarizations of all agents (Figure 6.9 C) are normally distributed around a rather high value (about 1.5, which corresponds to a summed vector of magnitude 0.3; in other words, the agents are about 30% polarized). However, when agents on excursions are excluded from the analysis, as was done in Chapter 4 for the empirical data (Figure 4.1 D), a bimodal distribution of polarizations emerges, closely mimicking the empirical data distribution even in the values of the distribution peaks. Thus, the current model, despite containing no alignment force, displays the bimodal polarization distribution that was attributed above to the distinction between shoaling and schooling. It is not entirely clear how the model exhibits such remarkable behavior. Since agents that are on excursions are made more attractive to other agents, it is possible that a single agent on an excursion has several followers. If these followers happened to approach the agent they were all following from the same direction their bearings would, of

\textsuperscript{14} Though no exploration of varying parameter values is presented, repeated unsuccessful attempts were made to improve the similarity of the model to the empirical data by manipulating both $P_E$ and $P_C$. 
necessity, align. Thus, the agent on an excursion emerges as an unwitting leader of the shoal (or, more correctly, of the school). On other occasions, when no single agent has a large enough following, the bearings of the agents are mostly uncorrelated and give rise to the ‘shoaling’ peak of the distribution. This raises the intriguing possibility that a similar mechanism is responsible for episodes of schooling in shoals of real fish. In other words, the model data suggest that the leader of a shoal (in the model, an agent attempting to depart the shoal) follows a different set of rules than the remainder of the fish in the shoal. This is the same conclusion we were led to by the comparison of the empirical data to Couzin et al.’s (2002) model above. Much more work is required, both theoretical and empirical, to examine the possibility that this or a similar mechanism account for the emergence of leaders in shoals.

Finally, Figure 6.9 E shows the distribution of bearing to nearest neighbor for the model data. As expected, the distribution is uniform ($p = 0.995$). Though the rules of the model require that agents orient towards another agent – unless they are on an excursion – the agents being followed are selected on the basis of their attractiveness, which is a function of whether or not they are on an excursion, and not based on their proximity to the focal agent.

Thus, the current model, whilst failing, for the most part, to generate cohesive shoals of a density approaching that of real shoals, nonetheless exhibits several behaviors that real shoals do but which existing models rarely capture. Combining the current model with a more detailed model of shoaling, such as either of the two models presented above, may result in a more detailed explanation of shoaling than is currently available and may point the way to possible mechanisms of shoaling that could then be explored experimentally.
6.4 Discussion

The direct comparison of model predictions to empirical data is informative not only on how well each model fits the data but also on possible mechanisms underlying the structure of the data. Comparisons to both models implied that the behavior of the zebrafish on the fifth day of the experiment was more variable than their behavior on the first day, and thus less amenable to prediction by any set of rules. Furthermore, that one set of rules – Couzin et al.’s – predicted movement at the rear of the shoal better than it did movement at its center suggests quite strongly that the behavior of zebrafish at different positions in the shoal is qualitatively different. Further work is required to determine what, exactly, these differences are.

The current analysis also permitted an indirect comparison between the different models. Overall, when each model was tested with its best-fit parameters (assuming the ideal values were not hiding in an unexplored region of parameter space) against the experimental data from day 1, Viscido et al.’s model had a narrower error distribution (Figure 6.7, yellow line) than Couzin et al.’s (Figure 6.4, yellow line), implying that the former model explained zebrafish behavior slightly better than the latter.

Both models struggled equally with data from later days of the experiment but this served mainly to emphasize how well they explained the data from days 1 and 2. When tested with the correct parameter values, predictions by both models rarely deviated by more than 50° from the direction actually taken by the fish. Given the simplicity of both models (relative to some of their competitors) and that the comparison of model to data relied solely on the positions and headings of the fish, such a level of specificity is remarkable. Thus, the most interesting finding, in my opinion, to emerge from these comparisons is that the behavior of shoals of zebrafish in a novel
environment (i.e. primarily on day 1 of the experiment) is, on average, predictable within quite a narrow range of values by relatively uncomplicated algorithms.

Finally, a novel movement rule was presented which, even on its own, reproduced some shoaling behaviors that existing models do not. There are several possible reasons why published models do not display excursions. First, modelers have primarily been interested in the mechanisms that hold a group together, rather than those that break it apart. Second, those authors that have compared or designed their models around empirical datasets have mostly used data from fish that shoal more tightly than zebrafish (e.g. golden shiners) and are therefore less likely to provide excursion data for simulation. Finally, the criteria for determining when a fish is or is not a member of a shoal have been underdeveloped (and those criteria that do exist underused) making excursions impossible to identify even when they did occur.
Figure captions

*Figure 6.1.* The zones in Couzin et al.’s (2002) model. zor – zone of repulsion; zoo – zone of orientation; zoa – zone of attraction; $\alpha$ – field of perception. Taken from Couzin et al. (2002, Figure 1), with permission.

*Figure 6.2.* Viscido et al.’s (2004) AAR function. $F_{\text{max}}$ – maximal force the agent can be subjected to; $\delta_r$ – beginning of repulsion zone; $\delta_p$ – preferred NND; $\delta_a$ – beginning of attraction zone; $\delta_m$ – maximal attraction distance; the height of the shaded rectangle represents the strength of the alignment force, $A_{ij}$. Taken from Viscido et al. (2004, Figure 1b), with permission.

*Figure 6.3.* Error distributions for Couzin et al.’s model for different values of the radius of $zoo$. Distributions for values above 15 cm are not visible as they are covered by the 50 cm distribution. See text for details.

*Figure 6.4.* Error distributions by day for Couzin et al.’s model.

*Figure 6.5.* Error distributions by rank for Couzin et al.’s model. $r$ – rank (see text for details).

*Figure 6.6.* Error distributions for Viscido et al.’s model for different values of $\delta_m$.

*Figure 6.7.* Error distributions by day for Viscido et al.’s model.

*Figure 6.8.* Error distributions by rank for Viscido et al.’s model. $r$ – rank (see text for details).

*Figure 6.9.* Model suggestion measure distributions. A – NND; B – IID; C – Polarization including all fish; D – Polarization excluding agents on excursions; E – Bearing to nearest neighbor. See text for details.
Figure 6.1
Figure 6.2

[Graph showing the relationship between attraction/repulsion force $S_{ij}$ and the distance between fish $i$ and $j$.]
Figure 6.3
Figure 6.4
Figure 6.5

![Graph showing distribution of error degrees for different values of r:]

- $r = 0$
- $0 < r \leq 0.25$
- $0.25 < r \leq 0.5$
- $0.5 < r \leq 0.75$
- $0.75 < r < 1$
- $r = 1$

Error (degrees)
Figure 6.6
Figure 6.7

![Graph showing error distribution over different days.](image-url)
Figure 6.8

![Graph showing error distribution with different ranges of r values.]

- $r = 0$
- $0 < r \leq 0.25$
- $0.25 < r \leq 0.5$
- $0.5 < r \leq 0.75$
- $0.75 < r < 1$
- $r = 1$

Error (degrees)
Figure 6.9
Chapter 7: Summary and Conclusions

The preceding chapters present a detailed account of zebrafish shoaling under a variety of conditions and the overall picture that emerges from all the data is of a flexible, dynamic behavior. The manner in which zebrafish shoal is sensitive to habituation, the threat of predation, the possibility of food, and even the expected presence of food. Shoaling differs between different populations of zebrafish and between shoals of different sizes. Many of the characteristics of shoaling explored here are shared by all groups of animals and a better understanding of zebrafish shoaling will enhance our knowledge of the general principles of collective motion.

None of these effects could have been brought to light without a robust and detailed description of shoaling, incorporating several different measures combined across multiple time-scales. Developing a better description of shoaling is the first step in studying the mechanisms – physiological or genetic – that drive it. As the data presented here show, this description permits the detection of even subtle changes in shoaling in response to environmental manipulations.

Some of the measures explored were found to be insensitive gauges of the nature of shoaling. The number of excursions, for example, rarely tracked the effects of environmental manipulations. However, the mean duration of excursions, a closely related value, was revealed as an excellent measure that clearly differentiated between different episodes of shoaling.

Figure 7.1 presents one method of displaying this multi-dimensional data in an easy-to-read graph. Each so-called ‘spider-web’ graph characterizes shoaling for one strain from Experiment 1. Each radial line emanating from the center of the graph represents an axis, a single measure. The distance along the axis (from the center) of the corresponding polygon vertex represents the
value of the data for that measure. All measures were standardized to vary between 0 and 1 before being plotted. All the measures discussed in Chapters 3 and 4 (with the exception of bearing to nearest neighbor) are plotted on each graph (see figure captions). The resulting polygons enable easy (though perhaps not precise) comparison between, in this case, the shoaling of different strains of zebrafish. Figure 7.2 presents similar graphs for the SF strain of fish by day for Experiment 1, showing the complex progression of habituation.

As the graphs demonstrate, shoaling in zebrafish may be characterized as follows: shoal members typically swim at a speed of about 3.5 (± 2.5) cm/s, about 1 BL/s. Fish maintain a distance of about 3 (± 2.5) cm from their nearest neighbor and, in shoals of 8 fish, are usually at an average distance (IID) of about 8.5 (± 7) cm from their shoal-mates. When schooling, groups are faster and have a summed movement vector of around 0.98 (and can colloquially be said to be 98 % polarized). Conversely, when shoaling, groups are typically slower and are around 60 % polarized. As they swim, fish continually peel away from and later rejoin the group, either individually or in sub-groups. Fish leave the group from the back by speeding up and turning away from the center of the group. Solitary individuals stay away longer than small sub-groups before rejoining the main shoal.

As fish habituate to their environment, achieved experimentally either through extending the length of the session or by repeated exposures to the tank on subsequent days, some characteristics of their shoals gradually change. Shoals become looser – NND and IID increase – and the speed of the shoal decreases (and, incidentally, becomes less variable). Shoaling gradually replaces schooling as the primary occupation of the group and the polarization of the group while shoaling decreases, though that of the remaining schooling episodes does not. Excursions away from the shoal become more frequent and of longer duration. Eventually the
shoal begins to dissolve into smaller sub-groups and, if the experiment were sufficiently extended, might eventually disappear altogether (as reported, for example, by Delaney et al., 2002).

Different strains of zebrafish shoal somewhat differently. This immediately suggests that there may be a genetic component to the variation in shoaling characteristics, though it should be noted that the majority of the differences were observed between the most physically divergent strain, the LF, and the other two strains. LF zebrafish were slower and their shoals less dense (NND and IID are higher) than those of SF or AB fish.

The above is, to the best of my knowledge, the most detailed account of the shoaling of any species reported to date. This account should be useful to modelers of shoaling as it can be used to calibrate models and set ecological limitations on their parameter values, as demonstrated in Chapter 6. In addition, this account could be used as a form of standardized behavioral assay for examining the effects of genetic or pharmacological manipulations on zebrafish shoaling. Using the data above as a null distribution, it should be possible to detect shoals that are ‘misbehaving’ as a result of some manipulation. We have begun to conduct such studies using pharmacological manipulations (specifically, exposing zebrafish to various concentration of alcohol, nicotine, and SCH23390, a dopamine D2R antagonist). Finally, the methods developed here could be used to examine the shoaling of other species in greater detail than has previously been reported and would allow comparisons to be made between species, revealing similarities and species-specific differences in shoaling (or even flocking and herding).

Moving in groups has been suggested to have broadly similar functions in most group-living species (Krause & Ruxton, 2002). Shoaling, flocking, or herding all provide protection from predation and enhance foraging success. As such, the mechanisms by which these adaptive
advantages are manifested are similar in some respects (and different in others) across many species. For instance, a well defined narrow range of NND values seems to characterize all shoals and flocks examined to date (see Chapter 1) and this likely also holds true for herds of mammals. Detailed exploration of how shoaling varies under different internal and external (environmental) conditions, as presented here, will allow not only informative comparisons amongst species and between genera but also, gradually, the discovery of the basic rules of collective motion common to most or all animal aggregations. It is worth remembering that humans, too, are a group-living species.

Finally, the relationship of shoaling to cognition remains to be discussed. The data presented here demonstrate that shoaling is a flexible, adaptable behavior. Zebrafish alter their interactions with their shoal-mates when the number of fish present changes, when a predator appears, or when food is presented or is expected to be presented. Shoaling is modulated by both learning and habituation. Thus, as a functional category of behavior, shoaling reveals the extent of some of the social cognitive skills employed by fish.

Social cognition, as intended here, must be clearly delimited from social learning, which I define (following, e.g., Shettleworth, 2010) as learning that derives from social sources of information – such as the behavior of conspecifics – independent of the content of the information. Thus, learning where food is to be found by watching the joyful dance of a sated conspecific is social learning but the knowledge thus acquired, about the location of food, and the application of that knowledge do not constitute social cognition. Rather, the type of social cognition referred to here involves information that is inherently social in nature, such as the locations of conspecifics, their current motivational or nutritional or oestral state, or information on relative dominance (possibly even third-party discriminations not involving the subject itself; such feats have been
demonstrated in some fish species; e.g. Oliveira et al., 1998). It remains unknown, for example, how many conspecifics shoaling fish attend to, whether this varies between shoals and schools, whether they attend to conspecifics at all times (or do they, for instance, sometimes ignore the rest of the shoal as suggested by the behavioral rule governing the model constructed in Chapter 6), and whether the degree of influence of conspecifics depends on their proximity to the subject or on some other factor (such as preferred destination, for example; Partridge, 1981; Couzin et al., 2005). In addition, the effects of learning on shoaling behavior, as demonstrated in Experiment 4 of the present work, have not been studied in detail (Brown & Laland, 2003).

Shoaling as an indicator of social cognition has been little considered. Social cognition is well studied in primates and birds (see Shettleworth, 2010) and has also been discussed in the context of social insects (e.g. Sumpter, 2006) – though usually as a prelude to detailed mechanistic explanations that seemingly preclude what most of us would consider cognition. However, knowledge of and about conspecifics and the rules that govern interactions with them in the species between these two extremes – such as fish – have primarily been the preserve of behavioral ecologists and ethologists, who bring to the field (literally) a different set of questions and preconceptions than experimental psychologists. The importance of using shoaling to explore cognition rests precisely on this point, that the types of questions psychologists will ask, and the types of answers they will seek, are different from those that have been asked of the behavior to date. For instance, little attention has been paid to comparative studies of shoaling and to relating the characteristics of a species’ shoaling to its ecology. To the best of my knowledge, no-one, prior to the current work, has sought an empirical basis for the distinction between shoaling and schooling, as presented in Chapter 4, or asked how general this distinction is (e.g., do birds have more than one mode of flocking?). The examination of these questions will, I hope, greatly expand our knowledge about and understanding of shoaling and other forms
of collective motion.
Figure captions

*Figure 7.1.* Spider-web graphs of all shoaling measures by strain for Experiment 1. Each line radiating from the center of the web represents an axis. The extension of the polygon along each spoke represents the value of the data on that axis. All measures are standardized before being plotted, such that each axis runs from 0 (at the center of the web) to 1 (at the edge). The measures, clockwise from the top, are: IID, NND, Speed, SGS, Exc Dur – excursion duration, Num Exc – number of excursions, Flexion – polarization level of the inflection point between the schooling and shoaling distributions, School – mode of the schooling polarization distribution component, Shoal – mode of the shoaling polarization distribution component. A – AB strain; B – LF strain; C – SF strain.

*Figure 7.2.* Spider-web graphs of all shoaling measures by day, for SF fish in Experiment 1. All axes as in Figure 7.1. A – Day 1; B – Day 2; C – Day 3; D – Day 4; E – Day 5.
Figure 7.1
Figure 7.2
References


Chen, J., Patel, R., Friedman, T.C., & Jones, K.S. (2010). The behavioral and pharmacological actions of NMDA receptor antagonism are conserved in zebrafish larvae. *International Journal of Comparative Psychology, 23,* 82-90


Appendix A: Equations and Formulae

A.1 Formulae for measures

*Nearest Neighbor Distance (NND).* For a particular frame, the NND gives for each individual $i$ the linear distance to its nearest neighbor. The mean value for each frame is reported.

\[
NND_i = \text{Min}(D_{ij}), \quad \text{where} \quad D_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}.
\]

When occlusions occur, several fish may be assigned the same position leading to NND values of 0. Such cases were excluded from all analyses (no NND value was given).

*Inter-Individual Distance (IID).* For a particular frame, the IID gives for each individual $i$ the mean distance to all other fish. The mean value for each frame is reported. \[IID_i = \frac{1}{N-1} \sum_{j \neq i} D_{ij},\]

where \[D_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \]

*Speed.* For a frame $T$, the momentary speed of every individual fish $i$ is given as:

\[
S_i = \sqrt{[x_i(T) - x_i(T + \Delta t)]^2 + [y_i(T) - y_i(T + \Delta t)]^2}, \quad \text{where} \quad \Delta t \text{ is the resolution at which the speed is measured, and is usually taken to be 0.5 sec. Note that the fish is assumed to have moved in a perfectly straight line between the two points. The mean value for each frame is reported.}

*Polarization.* Different authors have reported different measures of the polarization of a shoal. We chose to measure the circular standard deviation of the headings of the fish. The circular standard deviation is given by \[C = \sqrt{-2 \cdot \log(|R|)}, \quad \text{where} \quad |R| \text{ is the absolute size of the summed vector of all the fish. When all the fish are pointing in the same direction, } |R| \rightarrow 1 \text{ and } C \rightarrow 0; \text{ when all fish are pointing in different directions, } |R| \rightarrow 0 \text{ and } C \rightarrow \infty. \text{ Thus, } C \text{ is actually} \]
the inverse of the polarization, and is greater when the shoal is less polarized. Hemelrijk & Kunz (2004) reported a similar measure and called it ‘confusion’ to express its inverted relationship with orderliness.

A.2 Other mathematical formulae used

**Kernel Density Estimator (KDE).** The KDE evaluates the density distribution of a list \{A\} at any value \(x\). A Gaussian kernel is used. The density is given by:

\[
\hat{f}_h(x) = \frac{1}{Nh} \sum_{i=1}^{N} \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-A_i)^2}{2h^2}},
\]

where the subscript \(h\) indicates that the value of \(f\) depends on the bandwidth chosen, and where \(N\) is the length of the list of data \{A\}. For a more detailed definition and guidelines for selecting a bandwidth, see Chatfield (2002). For angular measures (such as bearing to a nearest neighbor) a circular KDE is used, in which \(x\) is replaced with \([(x + 180) \mod 360] - 180\), thus wrapping \(f(x)\) around a circle.

**Kolmogorov-Smirnov (KS) Test.** The KS test compares 2 distributions (either 2 data distributions or one data and one theoretical distribution). The test statistic, \(D\), is given by \(D = \text{Max} |S_{N1}(x) - S_{N2}(x)|\), where \(S_{N1}(x)\) and \(S_{N2}(x)\) are the cumulative density distributions (CDF) of the two datasets, respectively. Thus, \(D\) is the maximal distance between the plots of the two CDF functions. The significance of \(D\) is given by \(P(D > d) = Q_{KS}(d[\sqrt{N_e} + 0.12 + \frac{0.11}{\sqrt{N_e}}])\), where \(d\) is the observed value of the KS statistic, \(N_e\) is the effective sample size, and

\[
Q_{KS}(\lambda) = 2 \sum_{j=1}^{\infty} (-1)^{j-1} e^{-2j^2\lambda^2}.
\]

For 1-way tests, \(N_e\) is the number of samples in the dataset; for 2-way comparisons, \(N_e = \frac{n_1 * n_2}{n_1 + n_2}\), where \(n_1\) and \(n_2\) are the sample sizes of the two datasets.
(Chatfield, 2002). For most of the analyses in the present work, $N_c$ was divided by the number of frames per second tracked (usually 12) to overcome the non-independence of consecutive measurements (see Chapter 2).

**Periodogram.** The periodogram is a Fourier power spectrum of a discretely sampled time-series. It expresses the power of each harmonic in the data as a function of the period of the harmonic. The periodogram of a time-series $\{X(t)\}$ at period $\omega$ is given by

$$I(\omega) = \frac{1}{N} \left[ \left( \sum_{t=0}^{T-1} X(t) \cdot \cos(\omega t) \right)^2 + \left( \sum_{t=0}^{T-1} X(t) \cdot \sin(\omega t) \right)^2 \right],$$

where $N$ is the sample size (Frescura et al., 2007).

**Lomb-Scargle Test.** The Lomb-Scargle test evaluates the significance of peaks in a periodogram. The following formulae for the test are taken from Hernandez (1999; see also Frescura et al., 2007). As these papers are relatively unknown, the full equations are given here. If the value (power) of the periodogram at period $p$ is $I_p$, then the test statistic $\gamma = \frac{I_p}{2\sigma_Y^2}$, where $\sigma_Y^2$ is the variance of the time-series data (by Parseval’s theorem, variance is the same in the time and frequency domains). Hernandez (1999) shows that $P(\gamma > z) = 1 - (1 - e^{-\frac{z}{2}})^\frac{N}{2}$, where $N$ is the sample size and $z$ is any value to be exceeded. Thus, a significant peak is one for which

$$\gamma > -2 * \log(1 - P_0^n)$$

where $n = \frac{N}{2}$, $N$ is the sample size, and $P_0$ is the required significance level (e.g. 0.05). Note that the value of $\gamma$ depends on the variance of the data (and not only on the power of the individual periodogram peak) and thus the critical value for significance will vary between datasets.
**Likelihood.** The Likelihood, $L$, of a set of parameters $\theta$ is a measure of the probability that a particular set of data $\{x\}_n$ were drawn from the model described by $\theta$. $L(\theta \mid \{x\}_n) = \prod_{i=1}^{n} f(x_i \mid \theta)$, where $n$ is the number of data-points and $f(x_i \mid \theta)$ is the value (probability) of the model at $x_i$. $L$ depends on the number of data-points and is thus best used to compare the fits of two different models to the same dataset. Higher likelihood values represent better fits between model and data.

**Bayesian Information Criterion (BIC):** The BIC is a measure used for selecting between models with different numbers of parameters. The measure penalizes the model for each added predictor slightly more than the similar Akaike Information Criterion (AIC). $BIC = -2\log(L) + k\log(n)$, where $L$ is the likelihood score of the model (see above), $n$ is the number of data-points, and $k$ is the number of predictors in the model. The data are more likely to conform to models with a lower BIC. In the present case, due to the noisiness of the data, simpler models were preferred where the improvement in BIC that resulted from the addition of a model element was less than 1% of the BIC.
## Appendix B: Additional Figures and Tables

### B.1 Tables

Table B.1. *Results of KS tests on excursion duration by day in Experiment 1.* (See section 3.3.2).

Comparisons between excursion duration distributions for each day of the experiment are shown, for each strain separately. D – KS statistic; p – significance values of the KS statistic. µ – mode of the excursion duration distribution; σ – standard deviation of the excursion duration distribution. The columns under D and p refer to pairwise comparisons by day. Thus, the first cell under D3 compares day 3 to day 1 excursions; the second cell under D3 compares day 3 to day 2 excursions, and so on. Only one half of the KS comparison table is filled in, for clarity. Non-significant p values are highlighted.

<table>
<thead>
<tr>
<th>Strain</th>
<th>Day</th>
<th>µ</th>
<th>σ</th>
<th>D</th>
<th>p</th>
<th>D</th>
<th>p</th>
<th>D</th>
<th>p</th>
<th>D</th>
<th>p</th>
<th>D</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>D2</td>
<td>D3</td>
<td>D4</td>
<td>D5</td>
<td>D2</td>
<td>D3</td>
<td>D4</td>
<td>D5</td>
<td>D2</td>
<td>D3</td>
</tr>
<tr>
<td>AB</td>
<td>D1</td>
<td>3.08</td>
<td>2.66</td>
<td>0.043</td>
<td>0.166</td>
<td>0.132</td>
<td>0.290</td>
<td>0.720</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
<td></td>
</tr>
<tr>
<td></td>
<td>D2</td>
<td>3.20</td>
<td>2.48</td>
<td>0.178</td>
<td>0.143</td>
<td>0.301</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>D3</td>
<td>4.28</td>
<td>3.72</td>
<td>0.041</td>
<td>0.124</td>
<td></td>
<td>0.500</td>
<td>&lt; 0.001</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>D4</td>
<td>3.79</td>
<td>3.42</td>
<td>0.158</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>D5</td>
<td>5.71</td>
<td>4.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>D2</td>
<td>D3</td>
<td>D4</td>
<td>D5</td>
<td>D2</td>
<td>D3</td>
<td>D4</td>
<td>D5</td>
<td>D2</td>
<td>D3</td>
</tr>
<tr>
<td>LF</td>
<td>D1</td>
<td>3.07</td>
<td>2.77</td>
<td>0.166</td>
<td>0.256</td>
<td>0.201</td>
<td>0.316</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>D2</td>
<td>4.02</td>
<td>3.84</td>
<td>0.110</td>
<td>0.057</td>
<td>0.186</td>
<td>&lt; 0.001</td>
<td>0.174</td>
<td>&lt; 0.001</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>D3</td>
<td>4.76</td>
<td>4.92</td>
<td>0.064</td>
<td>0.106</td>
<td></td>
<td>0.030</td>
<td>&lt; 0.001</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>D4</td>
<td>4.50</td>
<td>4.50</td>
<td>0.153</td>
<td></td>
<td></td>
<td></td>
<td>&lt; 0.001</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>D5</td>
<td>5.03</td>
<td>6.80</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SF</td>
<td>D1</td>
<td>2.23</td>
<td>1.59</td>
<td>0.205</td>
<td>0.457</td>
<td>0.388</td>
<td>0.494</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>D2</td>
<td>2.99</td>
<td>2.27</td>
<td>0.281</td>
<td>0.215</td>
<td>0.326</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>D3</td>
<td>4.73</td>
<td>3.54</td>
<td>0.071</td>
<td>0.054</td>
<td></td>
<td>0.002</td>
<td>0.043</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>D4</td>
<td>4.09</td>
<td>3.36</td>
<td></td>
<td>0.111</td>
<td></td>
<td></td>
<td>&lt; 0.001</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>D5</td>
<td>5.00</td>
<td>4.25</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table B.2 *Results of KS tests on excursion type by excursion duration in Experiment 1.* (See section 3.3.2). Comparisons between excursion duration distributions by excursion type are shown, for each strain separately. T1 – excursions of type 1; T2 – excursions of type 2, and so on. All other symbols as in Table B.1. Non-significant p values are highlighted.

<table>
<thead>
<tr>
<th>Strain</th>
<th>Type</th>
<th>μ</th>
<th>σ</th>
<th>D</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>T2</td>
<td>T3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AB</td>
<td>T1</td>
<td>5.05</td>
<td>5.09</td>
<td>0.162</td>
<td>0.185</td>
</tr>
<tr>
<td></td>
<td>T2</td>
<td>4.12</td>
<td>3.45</td>
<td>0.026</td>
<td>0.038</td>
</tr>
<tr>
<td></td>
<td>T3</td>
<td>4.06</td>
<td>3.28</td>
<td>0.054</td>
<td>0.071</td>
</tr>
<tr>
<td></td>
<td>T4</td>
<td>4.46</td>
<td>3.33</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LF</td>
<td>T1</td>
<td>5.04</td>
<td>5.47</td>
<td>0.103</td>
<td>0.153</td>
</tr>
<tr>
<td></td>
<td>T2</td>
<td>4.04</td>
<td>4.39</td>
<td>0.062</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td>T3</td>
<td>3.72</td>
<td>3.85</td>
<td>0.09</td>
<td></td>
</tr>
<tr>
<td></td>
<td>T4</td>
<td>4.19</td>
<td>5.01</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SF</td>
<td>T1</td>
<td>4.92</td>
<td>4.78</td>
<td>0.167</td>
<td>0.178</td>
</tr>
<tr>
<td></td>
<td>T2</td>
<td>3.90</td>
<td>3.26</td>
<td>0.022</td>
<td>0.015</td>
</tr>
<tr>
<td></td>
<td>T3</td>
<td>3.75</td>
<td>3.18</td>
<td>0.022</td>
<td></td>
</tr>
<tr>
<td></td>
<td>T4</td>
<td>3.98</td>
<td>3.14</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table B.3 *Results of KS tests on SGS by testing day for Experiment 1.* (See Section 3.3.3).

Comparisons between SGS distributions by testing day are shown, for each strain separately. All symbols as in Table B.1. Non-significant p values are highlighted.

<table>
<thead>
<tr>
<th>Strain</th>
<th>Day</th>
<th>µ</th>
<th>σ</th>
<th>D2</th>
<th>D3</th>
<th>D4</th>
<th>D5</th>
<th>D2</th>
<th>D3</th>
<th>D4</th>
<th>D5</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AB</td>
<td>D1</td>
<td>0.36</td>
<td>1.36</td>
<td>0.031</td>
<td>0.100</td>
<td>0.055</td>
<td>0.306</td>
<td>0.564</td>
<td>&lt; 0.001</td>
<td>0.043</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td></td>
<td>D2</td>
<td>0.21</td>
<td>1.28</td>
<td></td>
<td>0.081</td>
<td>0.066</td>
<td>0.300</td>
<td>&lt; 0.001</td>
<td>0.008</td>
<td>&lt; 0.001</td>
<td></td>
</tr>
<tr>
<td></td>
<td>D3</td>
<td>0.05</td>
<td>1.46</td>
<td></td>
<td>0.078</td>
<td>0.232</td>
<td></td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
<td></td>
</tr>
<tr>
<td></td>
<td>D4</td>
<td>0.40</td>
<td>1.56</td>
<td></td>
<td></td>
<td></td>
<td>0.251</td>
<td></td>
<td>&lt; 0.001</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>D5</td>
<td>2.14</td>
<td>3.30</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LF</td>
<td>D1</td>
<td>0.46</td>
<td>1.40</td>
<td>0.042</td>
<td>0.157</td>
<td>0.137</td>
<td>0.207</td>
<td>0.223</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td></td>
<td>D2</td>
<td>0.47</td>
<td>1.55</td>
<td></td>
<td>0.116</td>
<td>0.095</td>
<td>0.166</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
<td></td>
</tr>
<tr>
<td></td>
<td>D3</td>
<td>0.83</td>
<td>2.19</td>
<td></td>
<td>0.027</td>
<td>0.054</td>
<td></td>
<td>0.765</td>
<td>0.053</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>D4</td>
<td>0.64</td>
<td>2.01</td>
<td></td>
<td></td>
<td></td>
<td>0.072</td>
<td></td>
<td>0.003</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>D5</td>
<td>0.85</td>
<td>2.48</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SF</td>
<td>D1</td>
<td>0.15</td>
<td>1.12</td>
<td>0.014</td>
<td>0.157</td>
<td>0.287</td>
<td>0.323</td>
<td>1.000</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td></td>
<td>D2</td>
<td>0.21</td>
<td>1.16</td>
<td></td>
<td>0.152</td>
<td>0.282</td>
<td>0.318</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
<td></td>
</tr>
<tr>
<td></td>
<td>D3</td>
<td>0.36</td>
<td>1.73</td>
<td></td>
<td>0.148</td>
<td>0.170</td>
<td></td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
<td></td>
</tr>
<tr>
<td></td>
<td>D4</td>
<td>0.93</td>
<td>2.95</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.245</td>
<td></td>
</tr>
<tr>
<td></td>
<td>D5</td>
<td>1.63</td>
<td>3.05</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table B.4 *Results of KS tests on distance measures by day for Experiment 1.* (See section 4.1).

Comparisons for IID and NND distributions between testing days are shown for each strain separately. All symbols as in Table B.1. Non-significant p values are highlighted.

<table>
<thead>
<tr>
<th>Meas.</th>
<th>Strain</th>
<th>Day</th>
<th>$\mu$</th>
<th>$\sigma$</th>
<th>D2</th>
<th>D3</th>
<th>D4</th>
<th>D5</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>D1</td>
<td>D2</td>
<td>D3</td>
<td>D4</td>
<td>D5</td>
</tr>
<tr>
<td>IID</td>
<td>AB</td>
<td>D1</td>
<td>6.99</td>
<td>3.98</td>
<td>0.023</td>
<td>0.063</td>
<td>0.071</td>
<td>0.255</td>
<td>0.889</td>
</tr>
<tr>
<td></td>
<td></td>
<td>D2</td>
<td>7.39</td>
<td>3.88</td>
<td>0.082</td>
<td>0.085</td>
<td>0.278</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td></td>
<td></td>
<td>D3</td>
<td>6.84</td>
<td>4.49</td>
<td>0.032</td>
<td>0.200</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td></td>
<td></td>
<td>D4</td>
<td>7.85</td>
<td>5.28</td>
<td>0.195</td>
<td></td>
<td></td>
<td></td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td></td>
<td></td>
<td>D5</td>
<td>21.44</td>
<td>16.39</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td></td>
<td>LF</td>
<td>D1</td>
<td>7.56</td>
<td>4.09</td>
<td>0.064</td>
<td>0.207</td>
<td>0.149</td>
<td>0.225</td>
<td>0.011</td>
</tr>
<tr>
<td></td>
<td></td>
<td>D2</td>
<td>8.52</td>
<td>5.24</td>
<td>0.145</td>
<td>0.088</td>
<td>0.169</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>D3</td>
<td>14.55</td>
<td>14.12</td>
<td>0.065</td>
<td>0.043</td>
<td></td>
<td></td>
<td>0.010</td>
</tr>
<tr>
<td></td>
<td></td>
<td>D4</td>
<td>10.95</td>
<td>10.89</td>
<td></td>
<td>0.108</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>D5</td>
<td>18.71</td>
<td>18.56</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>SF</td>
<td>D1</td>
<td>6.34</td>
<td>2.94</td>
<td>0.077</td>
<td>0.223</td>
<td>0.305</td>
<td>0.336</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td></td>
<td>D2</td>
<td>7.35</td>
<td>3.35</td>
<td>0.160</td>
<td>0.256</td>
<td>0.293</td>
<td></td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td></td>
<td></td>
<td>D3</td>
<td>9.27</td>
<td>6.34</td>
<td>0.106</td>
<td>0.145</td>
<td></td>
<td></td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td></td>
<td></td>
<td>D4</td>
<td>17.69</td>
<td>18.46</td>
<td></td>
<td>0.040</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>D5</td>
<td>22.63</td>
<td>18.88</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NND</td>
<td>AB</td>
<td>D1</td>
<td>2.86</td>
<td>1.74</td>
<td>0.055</td>
<td>0.055</td>
<td>0.109</td>
<td>0.473</td>
<td>0.045</td>
</tr>
<tr>
<td></td>
<td></td>
<td>D2</td>
<td>3.06</td>
<td>1.61</td>
<td>0.100</td>
<td>0.134</td>
<td>0.524</td>
<td></td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td></td>
<td></td>
<td>D3</td>
<td>3.04</td>
<td>1.89</td>
<td>0.054</td>
<td>0.424</td>
<td></td>
<td></td>
<td>0.050</td>
</tr>
<tr>
<td></td>
<td></td>
<td>D4</td>
<td>3.33</td>
<td>2.04</td>
<td></td>
<td>0.407</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>D5</td>
<td>6.83</td>
<td>5.40</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>LF</td>
<td>D1</td>
<td>3.10</td>
<td>1.69</td>
<td>0.107</td>
<td>0.423</td>
<td>0.273</td>
<td>0.413</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td></td>
<td></td>
<td>D2</td>
<td>3.45</td>
<td>1.91</td>
<td>0.330</td>
<td>0.178</td>
<td>0.318</td>
<td></td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td></td>
<td></td>
<td>D3</td>
<td>5.88</td>
<td>5.15</td>
<td>0.168</td>
<td>0.026</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>D4</td>
<td>4.01</td>
<td>3.14</td>
<td></td>
<td>0.155</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>D5</td>
<td>5.87</td>
<td>4.94</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>SF</td>
<td>D1</td>
<td>2.58</td>
<td>1.38</td>
<td>0.119</td>
<td>0.380</td>
<td>0.542</td>
<td>0.616</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td></td>
<td></td>
<td>D2</td>
<td>2.98</td>
<td>1.47</td>
<td>0.298</td>
<td>0.465</td>
<td>0.553</td>
<td></td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td></td>
<td></td>
<td>D3</td>
<td>3.85</td>
<td>2.52</td>
<td>0.192</td>
<td>0.274</td>
<td></td>
<td></td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td></td>
<td></td>
<td>D4</td>
<td>5.63</td>
<td>4.94</td>
<td></td>
<td>0.088</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>D5</td>
<td>6.92</td>
<td>5.20</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table B.5 *Results of KS tests on speed by day for Experiment 1.* (See section 4.1). Comparisons between speed distributions by testing day are shown for each strain separately. All symbols as in Table B.1. Non-significant p values are highlighted.

<table>
<thead>
<tr>
<th>Strain</th>
<th>Day</th>
<th>$\mu$</th>
<th>$\sigma$</th>
<th>$D_2$</th>
<th>$D_3$</th>
<th>$D_4$</th>
<th>$D_5$</th>
<th>$D_2$</th>
<th>$D_3$</th>
<th>$D_4$</th>
<th>$D_5$</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>AB</td>
<td>D1</td>
<td>7.20</td>
<td>4.15</td>
<td>0.125</td>
<td>0.459</td>
<td>0.292</td>
<td>0.503</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
<td></td>
</tr>
<tr>
<td></td>
<td>D2</td>
<td>6.61</td>
<td>3.56</td>
<td>0.403</td>
<td>0.227</td>
<td>0.448</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>D3</td>
<td>3.53</td>
<td>2.15</td>
<td>0.176</td>
<td>0.058</td>
<td>&lt; 0.001</td>
<td>0.030</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>D4</td>
<td>4.56</td>
<td>2.99</td>
<td></td>
<td>0.227</td>
<td></td>
<td>&lt; 0.001</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>D5</td>
<td>3.28</td>
<td>1.70</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LF</td>
<td>D1</td>
<td>3.16</td>
<td>2.06</td>
<td>0.036</td>
<td>0.162</td>
<td>0.139</td>
<td>0.244</td>
<td>0.393</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
<td></td>
</tr>
<tr>
<td></td>
<td>D2</td>
<td>3.10</td>
<td>2.38</td>
<td>0.144</td>
<td>0.121</td>
<td>0.226</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>D3</td>
<td>2.15</td>
<td>2.00</td>
<td>0.036</td>
<td>0.082</td>
<td></td>
<td>0.394</td>
<td>&lt; 0.001</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>D4</td>
<td>2.20</td>
<td>2.21</td>
<td></td>
<td>0.106</td>
<td></td>
<td></td>
<td>&lt; 0.001</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>D5</td>
<td>1.69</td>
<td>1.74</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SF</td>
<td>D1</td>
<td>6.61</td>
<td>4.24</td>
<td>0.274</td>
<td>0.372</td>
<td>0.342</td>
<td>0.335</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
<td></td>
</tr>
<tr>
<td></td>
<td>D2</td>
<td>4.71</td>
<td>2.46</td>
<td>0.192</td>
<td>0.106</td>
<td>0.212</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>D3</td>
<td>3.33</td>
<td>2.05</td>
<td>0.134</td>
<td>0.090</td>
<td></td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>D4</td>
<td>4.08</td>
<td>1.98</td>
<td></td>
<td>0.163</td>
<td></td>
<td></td>
<td></td>
<td>&lt; 0.001</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>D5</td>
<td>2.82</td>
<td>2.46</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table B.6  *Results of KS tests on polarization by day for Experiment 1.* (See section 4.1).

Comparisons between polarization distributions by testing day are shown for each strain separately. Note that polarization distributions are bimodal. $\mu_1$ – mode of the lower (schooling) component of the distribution; $\sigma_1$ – standard deviation of the lower (schooling) component; $\mu_2$ – mode of the upper (shoaling) component of the distribution; $\sigma_2$ – standard deviation of the upper (shoaling) component. All other symbols as in Table B.1. Non-significant p values are highlighted.

<table>
<thead>
<tr>
<th>Strain</th>
<th>Day</th>
<th>$\mu_1$</th>
<th>$\sigma_1$</th>
<th>$\mu_2$</th>
<th>$\sigma_2$</th>
<th>D2</th>
<th>D3</th>
<th>D4</th>
<th>D5</th>
<th>D2</th>
<th>D3</th>
<th>D4</th>
<th>D5</th>
</tr>
</thead>
<tbody>
<tr>
<td>AB</td>
<td>D1</td>
<td>0.17</td>
<td>0.13</td>
<td>0.61</td>
<td>0.57</td>
<td>0.030</td>
<td>0.252</td>
<td>0.126</td>
<td>0.205</td>
<td>0.638</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td></td>
<td>D2</td>
<td>0.19</td>
<td>0.12</td>
<td>0.53</td>
<td>0.77</td>
<td>0.245</td>
<td>0.118</td>
<td>0.197</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>D3</td>
<td>0.23</td>
<td>0.14</td>
<td>0.88</td>
<td>0.58</td>
<td>0.128</td>
<td>0.051</td>
<td>&lt; 0.001</td>
<td>0.077</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>D4</td>
<td>0.20</td>
<td>0.13</td>
<td>0.80</td>
<td>0.68</td>
<td>0.080</td>
<td>&lt; 0.001</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>D5</td>
<td>0.21</td>
<td>0.14</td>
<td>0.76</td>
<td>0.61</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LF</td>
<td>D1</td>
<td>0.26</td>
<td>0.14</td>
<td>1.02</td>
<td>0.55</td>
<td>0.132</td>
<td>0.077</td>
<td>0.066</td>
<td>0.146</td>
<td>&lt; 0.001</td>
<td>0.011</td>
<td>0.008</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td></td>
<td>D2</td>
<td>0.25</td>
<td>0.14</td>
<td>1.06</td>
<td>0.59</td>
<td>0.149</td>
<td>0.131</td>
<td>0.208</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>D3</td>
<td>0.22</td>
<td>0.15</td>
<td>1.20</td>
<td>0.54</td>
<td>0.019</td>
<td>0.069</td>
<td></td>
<td>0.981</td>
<td>0.005</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>D4</td>
<td>0.22</td>
<td>0.15</td>
<td>1.18</td>
<td>0.54</td>
<td>0.081</td>
<td></td>
<td>&lt; 0.001</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>D5</td>
<td>0.22</td>
<td>0.16</td>
<td>1.30</td>
<td>0.47</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SF</td>
<td>D1</td>
<td>0.21</td>
<td>0.13</td>
<td>0.60</td>
<td>0.52</td>
<td>0.085</td>
<td>0.238</td>
<td>0.128</td>
<td>0.248</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td></td>
<td>D2</td>
<td>0.24</td>
<td>0.13</td>
<td>0.69</td>
<td>0.56</td>
<td>0.171</td>
<td>0.069</td>
<td>0.193</td>
<td>&lt; 0.001</td>
<td>0.005</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
<td></td>
</tr>
<tr>
<td></td>
<td>D3</td>
<td>0.23</td>
<td>0.13</td>
<td>0.96</td>
<td>0.62</td>
<td>0.111</td>
<td>0.078</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>D4</td>
<td>0.23</td>
<td>0.14</td>
<td>0.84</td>
<td>0.58</td>
<td>0.124</td>
<td></td>
<td>&lt; 0.001</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>D5</td>
<td>0.16</td>
<td>0.12</td>
<td>1.15</td>
<td>0.54</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table B.7. Results of KS tests on all unimodal measures by day for Experiment 3. (See section 5.1.3). Comparisons between measure distributions by testing day are shown for each strain separately for excursion duration (top), NND (center), and speed (bottom). N – shoal size. All other symbols as in Table B.1. Non-significant p values are highlighted.

<table>
<thead>
<tr>
<th>Measure</th>
<th>N</th>
<th>Day</th>
<th>μ</th>
<th>σ</th>
<th>D</th>
<th></th>
<th></th>
<th>D</th>
<th></th>
<th></th>
<th>p</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>5</td>
<td>D1</td>
<td>2.23</td>
<td>1.44</td>
<td>0.238</td>
<td>0.385</td>
<td><strong>0.129</strong></td>
<td>&lt; 0.001</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>D2</td>
<td>3.16</td>
<td>2.06</td>
<td>0.148</td>
<td></td>
<td>0.479</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>D3</td>
<td>3.84</td>
<td>2.51</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Excursion</td>
<td>10</td>
<td>D1</td>
<td>2.00</td>
<td>1.81</td>
<td>0.263</td>
<td>0.042</td>
<td>&lt; 0.001</td>
<td></td>
<td><strong>0.993</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Duration</td>
<td></td>
<td>D2</td>
<td>3.06</td>
<td>2.90</td>
<td></td>
<td>0.232</td>
<td>&lt; 0.001</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>D3</td>
<td>1.96</td>
<td>1.92</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>D1</td>
<td>1.66</td>
<td>1.70</td>
<td>0.235</td>
<td>0.311</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>D2</td>
<td>2.85</td>
<td>2.14</td>
<td></td>
<td>0.077</td>
<td></td>
<td><strong>0.880</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>D3</td>
<td>3.16</td>
<td>2.21</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NND</td>
<td>5</td>
<td>D1</td>
<td>2.44</td>
<td>1.28</td>
<td>0.068</td>
<td>0.228</td>
<td><strong>0.107</strong></td>
<td>&lt; 0.001</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>D2</td>
<td>2.65</td>
<td>1.26</td>
<td></td>
<td>0.184</td>
<td>&lt; 0.001</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>D3</td>
<td>3.23</td>
<td>1.43</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>D1</td>
<td>2.67</td>
<td>1.65</td>
<td>0.041</td>
<td>0.163</td>
<td><strong>0.660</strong></td>
<td>&lt; 0.001</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>D2</td>
<td>2.87</td>
<td>1.66</td>
<td></td>
<td>0.194</td>
<td>&lt; 0.001</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>D3</td>
<td>2.43</td>
<td>1.27</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>D1</td>
<td>1.88</td>
<td>1.25</td>
<td>0.032</td>
<td>0.170</td>
<td><strong>0.897</strong></td>
<td>&lt; 0.001</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>D2</td>
<td>1.97</td>
<td>1.18</td>
<td></td>
<td>0.158</td>
<td>&lt; 0.001</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>D3</td>
<td>2.44</td>
<td>1.37</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Speed</td>
<td>5</td>
<td>D1</td>
<td>7.94</td>
<td>3.84</td>
<td>0.316</td>
<td>0.242</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>D2</td>
<td>4.27</td>
<td>3.57</td>
<td></td>
<td>0.157</td>
<td>&lt; 0.001</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>D3</td>
<td>6.10</td>
<td>3.31</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>D1</td>
<td>9.83</td>
<td>2.86</td>
<td>0.387</td>
<td>0.381</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>D2</td>
<td>6.78</td>
<td>3.05</td>
<td></td>
<td>0.029</td>
<td></td>
<td><strong>0.948</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>D3</td>
<td>7.01</td>
<td>2.97</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>D1</td>
<td>10.49</td>
<td>3.09</td>
<td>0.524</td>
<td>0.492</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>D2</td>
<td>5.08</td>
<td>3.35</td>
<td></td>
<td>0.094</td>
<td></td>
<td>0.008</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>D3</td>
<td>5.93</td>
<td>3.02</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table B.8: Results of KS tests on polarization distributions by day for Experiment 3. (See section 5.1.3). Comparisons between polarization distributions by testing day are shown for each strain separately. Note that some sessions have tri-modal distributions: three modes are reported only for sessions for which the distribution was best described by a three-component model (see text for details). N – shoal size. All other symbols as in Table B.6. Non-significant p values are highlighted.

<table>
<thead>
<tr>
<th>N</th>
<th>Day</th>
<th>µ1</th>
<th>σ1</th>
<th>µ2</th>
<th>σ2</th>
<th>µ3</th>
<th>σ3</th>
<th>D</th>
<th>p</th>
<th>D</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>1</td>
<td>0.13</td>
<td>0.12</td>
<td>0.47</td>
<td>0.72</td>
<td>--</td>
<td>--</td>
<td>0.162</td>
<td>0.096</td>
<td>&lt; 0.001</td>
<td>0.005</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.16</td>
<td>0.13</td>
<td>0.67</td>
<td>0.75</td>
<td>--</td>
<td>--</td>
<td>0.073</td>
<td></td>
<td></td>
<td>0.067</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.16</td>
<td>0.13</td>
<td>0.90</td>
<td>0.67</td>
<td>--</td>
<td>--</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>1</td>
<td>0.22</td>
<td>0.12</td>
<td>0.55</td>
<td>0.21</td>
<td>--</td>
<td>--</td>
<td>0.071</td>
<td>0.119</td>
<td>0.079</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.24</td>
<td>0.12</td>
<td>0.56</td>
<td>0.30</td>
<td>9.8</td>
<td>8.5</td>
<td>0.051</td>
<td></td>
<td></td>
<td>0.393</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.24</td>
<td>0.12</td>
<td>0.54</td>
<td>0.26</td>
<td>1.01</td>
<td>0.47</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>1</td>
<td>0.28</td>
<td>0.12</td>
<td>0.52</td>
<td>0.20</td>
<td>1.07</td>
<td>0.59</td>
<td>0.145</td>
<td>0.091</td>
<td>&lt; 0.001</td>
<td>0.010</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.41</td>
<td>0.20</td>
<td>1.22</td>
<td>0.56</td>
<td>--</td>
<td>--</td>
<td>0.098</td>
<td></td>
<td></td>
<td>0.005</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.45</td>
<td>0.23</td>
<td>1.05</td>
<td>0.48</td>
<td>--</td>
<td>--</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
B.2 Figures

Figure captions

Figure B.1. Experimental setup for Experiments 1, 2, 3, and 4. A – testing tank; B – one of the two fluorescent lights, positioned to prevent glare; C – video camera. The entire testing tank was visible in the frame of the video.

Figure B.2. Experimental setup for Experiment 5. A – testing tank; B – video camera; C – projector. The projector provided the only light to the tank. The entire testing tank was visible in the frame of the video.

Figure B.3. Density distributions of excursion durations by strain for Experiment 1. The LF distribution is significantly different from those of the other two strains.

Figure B.4. Density distributions of excursion durations by excursion type for Experiment 1. Note that excursions of type 1 – when a single individual leaves the shoal – are of longer duration than other types of excursions – when a sub-group splits from the shoal – which do not differ from each other.

Figure B.5. Density distributions of individual speed in the 1 sec before (blue) and the first sec of (red) an excursion in Experiment 1. The modes of the ‘before’ and ‘during’ distributions are 3.7 (± 3.8) and 4.12 (± 4.66) cm/sec respectively. The distributions are significantly different from each other (modified KS test, D = 0.08, p < 0.001).

Figure B.6. Density distributions of NND by day for LF fish in Experiment 1. Note the appearance of a second peak around 13 cm on days 4 and 5 (see Section 4.1 for details).
Figure B.7. Density distributions of periodic oscillations in shoaling measures by strain for

Experiment 1. Each panel displays the distribution of significant oscillations in one measure by the period of the oscillation (the inverse of the frequency). A – IID; B – NND; C – polarization; D – speed.
Figure B.1
Figure B.3
Figure B.4
Figure B.5
Figure B.7

A

B

C

D

Period (sec)

Period (sec)

Period (sec)

Period (sec)

P

P

P

P

0.0005

0.0005

0.0005

0.0005

0.001

0.001

0.001

0.001

0.0015

0.0015

0.0015

0.0015

0.002

0.002

0.002

0.002

0.0025

0.0025

0.0025

0.0025

AB
LF
SF
Copyright Acknowledgements

Figure 6.1. reprinted from Journal of Theoretical Biology, 218, Couzin, I.D., Krause, J., James, R., Ruxton, G.D., & Franks, N.R., Collective memory and spatial sorting in animal groups, pp. 1-11, Copyright 2002, with permission from Elsevier.

Figure 6.2 reprinted from Marine Ecology Progress Series, 273, Viscido, S.V., Parrish, J.K., & Grünbaum, D., Individual behavior and emergent properties of fish schools: a comparison of observation and theory, pp. 239-249, Copyright 2004, with permission from Inter-Research.